Title
Transcriptional Dynamics of Normal and Hand2-Deficient Cardiogenesis at Single-Cell Resolution

Author
de Soysa, Tarja Yvanka

Publication Date
2019

Peer reviewed|Thesis/dissertation

Transcriptional Dynamics of Normal and Hand2-deficient Cardiogenesis at Single-Cell Resolution

by
Tarja Yvanka De Soysa

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Benoit Bruneau
EFE9E009F6FE4D5...

Benoit Bruneau
Chair

DocuSigned by:
Deepak Srivastava

Deepak Srivastava

DocuSigned by:
Joseph Costello
917AF3FB58574BE...

Joseph Costello

Committee Members

I dedicate this dissertation to my loving husband and fellow graduate student, Tim.

Thank you for everything.

Acknowledgements

I am deeply grateful to my PhD advisor and science cheerleader, Deepak, for your support and mentorship throughout my graduate school career. Thank you for always being there when I needed you most, for the lessons you taught me about advocating for myself and being a good citizen in every community I'm a part of, and for the opportunity to be a part of the remarkable community of scientists in your lab.

Thank you to my thesis committee members for your guidance, encouragement and constructive feedback. Thank you, Benoit, for always taking the time to chat whenever I needed clarity in my research ideas. Thank you, Joe, for pushing me to think beyond my project and consider the broader implications of my work.

It has been a privilege to be a member of the Srivastava lab. I thank all my amazing lab mates for their kindness and support. Thank you for taking the time to listen to every practice talk, celebrating every successful litter dissection and for bringing my dream of choreographing a Hamilton number to fruition. You are a big reason why it's a joy to come into lab every day. Special shout out to Kathy Ivey and Joke Van Bemmel, for imparting your wisdom on being a rigorous experimentalist and helping me see that I have made tremendous progress as a scientist in the last 6 years. Huge thanks to Bethany Taylor, Karena Essex and Nicole Velasquez for making sure our lab is always running optimally and for always helping me procure a laser pointer. Thank you also to my fellow GICD scientists for your collaboration, mentorship and friendship.

The inclusive and supportive community at Gladstone is one that I am not sure I will be able to find again. Thank you to everyone in the Gladstone Cores, the Animal facility, Admin, Graphics and Communications who advance and elevate our science.

Thank you to Bob Mahley, for being my secondary mentor and for your advice about life and science. Your enthusiasm and joy of discovery is contagious and inspiring.

I came to UCSF because of the wonderful community that I witnessed as a recruit. I'm so grateful for the friendship and support of my fellow grad student friends and the administrative team of the Biomedical Sciences program. Being in this together with you, particular during preparation for quals, truly enriched my experience as a graduate student. Thank you to Scott Oakes, Arturo Alvarez-Buylla, Joanna Phillips, Daniel Lim, Takashi Mikawa, Matthew Springer, Sheng Ding, Rahul Deo, Brian Black, Bruce Conklin and Todd Mcdevitt for taking the time to mentor and advise me as I progressed through my studies in graduate school.

I have lived several continents away from my nuclear family in Sri Lanka for the past 12 years. I am truly grateful to my husband's family for taking me under their wing as a family member and helping me feel like the US is a place I can call home. To my parents, Sonali and Senaka, everything I do is to make you proud. Thank you for encouraging my educational goals even as they took me to the other side of the world, for teaching me that there is a lesson to be learned from every mistake and setback, and for reminding me to always keep moving forward. Thank you to my mother and grandmother, Dagma, for your prayers. To my sister Anouk, thanks for always being the 'chill' counterpoint to my 'stress machine' tendencies, and reminding me that 'it's just bacteria' every time I complained when my cloning failed.

I have been able to face and overcome countless daunting tasks and frustrating experimental challenges thanks to the mentorship and example of my Science Yodas, Jeeves and Casey. Thank you, Jeeves, for joining me on early morning litter dissections,

reading every fellowship application, and always having my back. Thank you, Casey, for challenging me to be a more thoughtful scientist, encouraging me to be confident and fearless in my pursuit of new skills, and helping me through many crises. I have learned so much from you both and am so grateful for your friendship and advice as I navigated my PhD journey.

I dedicate this dissertation to my husband Tim, without whom this adventure would have been much less bearable. Thank you for always reminding me that this was a learning experience and that I had enough grit to get through it, for making me laugh every day, celebrating every victory with me even in the face of your own challenges, and letting me cry on your shoulder during trying times. It has been a privilege to be in this together with you.

Contributions

Excerpts of the text from Chapters 2, 3 and 5 of this dissertation originally appeared in the following publication:

de Soysa, TY., Ranade, SS., Okawa, S., Ravichandran, S., Huang, Y., Salunga, HT., Schricker, A., del Sol, A., Gifford, CA & Srivastava, DS. Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. Nature (2019).

PMID: 31341279 doi: 10.1038/s41586-019-1414-x

Specific Contributions

TYdS, CAG and DS interpreted the data, and DS supervised the research that forms the basis for this dissertation. TYdS prepared chromium libraries, performed *in situ* hybridization experiments and analyzed data with Seurat and Monocle. TYdS and SSR dissected and processed embryos for single-cell library preparation and *in situ* hybridization experiments and conducted whole mount and section imaging. TYdS and AS performed genotyping of mice. SO, SR performed, and AdS conducted the computational modeling for cell fate determinant predictions. YH and HTS identified pregnant female mice by echocardiography.

Transcriptional Dynamics of Normal and *Hand2*-deficient Cardiogenesis at Single-cell Resolution

Tarja Yvanka de Soysa

Abstract

Organogenesis involves integration of myriad cell types, each progressing through successive stages of lineage specification, proliferation and differentiation. Establishment of unique gene networks within each cell drives fate determination and behavior, and mutations of the transcription factors that drive such networks can result in birth defects. Congenital heart malformations are the most common defects, and are caused by disruption of discrete subsets of progenitors that contribute to distinct cardiac structures. However, determining the transcriptional changes in individual cells that lead to organ-level defects in the heart has not been tractable. Here, we employed single-cell transcriptomics to interrogate early cardiac progenitor cells as they become specified during normal and abnormal cardiogenesis. We identified novel cell-type specific genes and uncovered additional heterogeneity within wild-type progenitor compartments. A network-based computational method that predicts lineage specifying transcription factors identified *Hand2* as a specifier of outflow tract cells but not right ventricular cells, despite failure of right ventricular formation in *Hand2*-null mice. Temporal single-cell transcriptome analysis of *Hand2*-null embryos revealed failure of outflow tract myocardium specification, whereas right ventricular myocardium was appropriately specified, but exhibited differentiation defects and failed to migrate into the developing heart. We found dysregulation of retinoic acid signaling that was associated with

posteriorization of anterior cardiac progenitors in *Hand2*-null mutant hearts and ectopic atrial gene expression in outflow tract cells and right ventricle precursors. This work reveals transcriptional determinants in individual cells that specify cardiac progenitor cell fate and differentiation, and exposes mechanisms of disrupted cardiac development and single-cell resolution, providing a framework to investigate congenital heart defects.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Cardiac and Embryo Terms

A – atria

AHF – anterior heart field

AVC – atrioventricular canal

bHLH – basic helix loop helix

CC – cardiac crescent

CHD – congenital heart defects

CPC – cardiac progenitor cell

EC – endothelial or endocardial cells

EMP – early myocardial progenitors

EndMT – endothelial to mesenchymal transition

FHF – first heart field

H – head

HE – hematoendothelial progenitors

HF – head fold

HT – heart tube

IVS – interventricular septum

LPM – lateral plate mesoderm

LV – left ventricle

MP – multipotent progenitors

NC – neural crest-derived mesenchyme

OFT – outflow tract

PA – pharyngeal arch

PEO – proepicardial organ

PM – paraxial mesoderm

pSHF – posterior second heart field

RA – retinoic acid

RARE-LacZ – retinoic acid response element-driven lacZ

RV – right ventricle

SHF – second heart field

SV – sinus venosus

WT – wild-type

Technologies and Computational terms

ATAC-seq – assay for transposase-accessible chromatin with sequencing

ChIP-seq – chromatin immunoprecipitation with sequencing

scATAC-seq – single-cell ATAC-seq

scRNA-seq – single-cell RNA sequencing

UMAP – uniform manifold approximation and projection

Chapter 1

Introduction

1.1 Congenital Heart Defects

Congenital heart defects (CHDs) are malformations of the heart and great vessels that arise due to disrupted embryonic cardiac development[1,2]. They encompass a wide spectrum of abnormalities that manifest in diverse cardiac structures and display a broad range in severity. CHDs are considered a major cause of fetal demise and are the most commonly occurring birth defect worldwide, affecting millions of newborns annually, with increasing global prevalence[3,4]. Recent advances in prenatal diagnosis, corrective surgical strategies and long-term care have resulted in over 75% of CHD patients who survive to adulthood[3]. While this represents a major clinical advancement, these patients remain at risk for late complications, such as cardiac arrhythmias, endocarditis, pulmonary hypertension, heart failure, neurodevelopmental deficits and other congenital abnormalities[1,5]. Thus, understanding how the molecular and cellular pathways underlying cardiac development are disrupted in these patients is critical for designing intervention and therapeutic strategies for the long-term sequelae of their CHDs[6].

Known non-genetic causes of CHD include environmental teratogens (i.e., pesticides and retinoic acid), maternal exposures (i.e., alcohol, maternal diabetes and phenylketonuria) and infectious agents (i.e., rubella)[1,3] and account for roughly 10% of CHD cases[2]. Defects in single genes account for 3-5% of CHD cases and include loss of function of core cardiac transcription factors such as NKX2.5[7], GATA4[8], MEF2C[9], TBX5[10] and TBX1[10]. Aneuploidies, copy number variations and *de novo* single nucleotide variants account for 33% of CHD cases, which include CHDs arising from malformation syndromes such as Down syndrome, trisomy 13, trisomy 18, Turner syndrome and DiGeorge syndrome[1,2,5]. The genetic underpinnings of over 50% of CHD cases remain

unknown. Collaborative efforts to map genetic variants associated with the disease are underway, aided by advances in sequencing and genomic technologies[5]. However, many of the genes in which CHD-linked mutations have been identified are novel, with unknown function and expression dynamics in cardiogenesis. Dissecting how such genes function in cardiac development in model organisms and with stem cell based cardiac differentiation systems is critical for understanding how their damaging variants could leads to CHDs.

1.2 Early mammalian cardiogenesis

A salient feature of most CHDs is that distinct structures of the heart are malformed, while the rest of the organ develops appropriately[11]. This reflects the integration of regionally and molecularly distinct cell types, that are specified from multiple progenitor domains, in assembling discrete structures of the embryonic heart. Indeed, embryonic cardiogenesis is an intricate process involving diverse cell type specification events, reciprocal cell-cell interactions and complex morphogenetic changes.

Work in model organisms such as the chick, frog, zebrafish and the mouse have revealed the molecular and cellular pathways underlying cardiogenesis[12]. The origins of the heart, which is the first organ to form in embryonic development, can be traced to the early gastrulating embryo. At this stage, which occurs at mouse embryonic day (E) 6.0, gastrulating mesoderm is rapidly sub-specified to committed *Mesp1*$^{+}$ precursors fated for specific anatomical locations in the heart[13,14]. These precursors leave the primitive streak at the dorsal region of the developing embryo and migrate towards the antero-lateral pole to form the lateral plate mesoderm (LPM)[15]. As they migrate and form the LPM, the

cardiac precursors are closely apposed to the anterior endoderm, which secretes signals such as Bone morphogenetic protein (Bmp)-2, Bmp-4 and Fibroblast growth factor (Fgf)-8 to induce their cardiogenic gene expression[12].

At E7.5 in the mouse, which corresponds to week 2 of human gestation, the first visible cardiac structure forms, resembling a crescent[6]. This cardiac crescent comprises progenitor cells described as the first heart field (FHF), with a second progenitor population called the second heart field (SHF) located medially and more posteriorly[16]. The FHF is a rapidly differentiating progenitor population that is marked by the potassium-channel subunit encoding gene, *Hcn4*[17], while SHF cells, defined by *Isl1* and *Tbx1* expression[18,19], remain in a prolonged progenitor-like state, likely because they are exposed to inhibitory Wnt signals emanating from the embryonic midline[6,20,21]. The SHF is also patterned along the anterior-posterior axis into two molecularly distinct domains that give rise to specific cardiac structures, the anterior heart field (AHF), that expresses genes such as *Fgf8* and *Fgf10*[6], and the posterior second heart field (pSHF), marked by *Nr2f2*, *Tbx5*, and *Wnt2*[16]. At E8.0, corresponding to week 3 in human development, the FHF cells migrate and fuse at the ventral midline to form a primitive heart tube, the presumptive left ventricle (LV), that begins to beat and consists of an inner layer of specialized cardiac endothelial cells, the endocardium, and an outer layer of myocardial cells. Between the two layers lies an extracellular matrix mainly composed of hyaluronan and chondroitin sulfates called the cardiac jelly[6,22]. Concomitantly, the SHF cells relocate dorsally into the pharyngeal mesoderm and begin to migrate into the heart at the arterial and venous poles, contributing to the majority of heart growth and chamber formation.

The AHF cells form the presumptive outflow tract (OFT) and right ventricle (RV) chamber, while the pSHF cells give rise to the atria and sinus venosus (SV)[16].

At E8.5, the heart breaks symmetry and undergoes rightward looping[23], to distinguish and align the RV and LV chambers. By E9.5, which is week 4 of human gestation, the heart comprises the two ventricles, a common atrial chamber, the atrioventricular canal (AVC), SV and OFT. At this stage, neuroectoderm-derived cardiac neural crest cells are migrating from the dorsal neural tube into the pharyngeal arches and the OFT at the arterial pole of the heart in response to OFT-derived guidance cue signals[24–26]. After they invade the OFT, these cells differentiate into smooth muscle cells and remodel the OFT to form the base of the pulmonary artery and the aorta; this enables blood from the ventricles to enter the systemic versus pulmonary circulation. Formation of the cardiac valves, which are affected in 20-30% of congenital heart defects[22,27], is also in progress at this developmental stage. To form these structures, the endocardial cells underlying the OFT and AVC undergo an endothelial-to-mesenchymal transition (EndMT) in response to *Bmp4* and *Bmp2* signals secreted by the overlying OFT and AVC myocardium, respectively. These prospective valve endothelial cells migrate into a bulge of cardiac jelly, called the endocardial cushions, to populate it with mesenchymal cells. Additionally, the transient proepicardial organ, comprising a population of cells that will begin to migrate over and sheathe the myocardium in a layer of epicardium, emerges at this stage[28]. A subset of these epicardial cells will later invade the underlying myocardium and give rise to cardiac fibroblasts and vascular smooth muscle cells[29].

Beyond E9.5, the heart continues to develop and remodel; this includes events such as trabeculation and compaction of the ventricular myocardium[30], development of the

sinoatrial node which consists of pacemaker cells that regulate cardiac contraction[31], remodeling and maturation of the endocardial cushions to form the cardiac valves[32] and septation of the OFT, ventricles and atria[33]. Thus, during the early stages of heart development (~E7.5-E9.5), a myriad cardiac cell subtypes are specified from their respective progenitor domains to contribute to distinct structures of the heart as it undergoes major morphogenetic changes. This multi-chambered organ then serves as a scaffold for further refinement, growth and maturation.

1.3 Patterning events during early cardiogenesis

Cardiac specification and morphogenesis are regulated by molecular patterning events that confer morphological asymmetry as well as regionalization of gene expression and cell fate. *Pitx2*-mediated left-right (L/R) asymmetry in the cardiac crescent and anterior-posterior patterning of the SHF by retinoic acid signaling are two major patterning events that occur in cardiogenesis.

The rightward looping of the heart at E8.5 is the first visible sign of asymmetric morphogenesis, but the molecular drivers that govern cardiac left-right identity are established much earlier. The most upstream signal that defines L/R identity at the cardiac crescent stage comes from the node, the signaling organizer of the gastrulating mouse embryo[34]. Cilia in the node exhibit left-sided rotational movement that causes a leftward movement of fluid, known as nodal flow[35]. This results in asymmetric expression and secretion of *Nodal*, a member of the *TGF-ß* family, in the left LPM. Through activation of its receptor complex (comprising its co-receptors, *Cfc1*, *Acvr1b* and *Acvr2b*) and subsequent downstream signaling, *Nodal* activates transcription of itself and its

antagonist *Lefty2* and the downstream transcriptional effector of asymmetry, *Pitx2*[23,36,37]. *Lefty-2* establishes a negative feedback loop that limits *Nodal* signaling to a short time window of just a few hours. Additionally, expression of a second *Nodal* antagonist *Lefty-1* in the embryonic midline is induced by *Shh* signals[38], to suppress *Nodal* expression in the right LPM. Once activated in the left LPM, asymmetric expression of *Pitx2* persists in the left side of the cardiac crescent and the SHF; a day later its expression persists on the left side of the heart tube, including the left prospective sinoatrial, ventricular and outflow tract region[39].

Once activated, *Pitx2* controls localized changes in cell migration, shape, extracellular matrix composition, regional proliferation and survival to drive asymmetric features of the developing heart[40,41]. It is important for the completion of heart looping (although its loss does not impact the direction of looping)[23], for asymmetric proliferation in the sinus venosus, where it represses proliferation on the left side, and for driving OFT lengthening by promoting myocardial proliferation in the left OFT[42]. Additionally, rotation of the OFT, which is critical for proper positioning of the aorta and pulmonary artery, is *Pitx2*-dependent[43]. *Pitx2* is also required for maintenance of left atrial identity as its loss leads to right atrial isomerism, where both atria have a right-specific gene expression pattern and morphology[44]. This condition is the most commonly occurring laterality defect-based CHD[45]. However, little is known about the transcriptional targets of *Pitx2* underlying establishment of L/R asymmetry. Moreover, the presence and identity of signals and transcriptional mediators that define right-sided identity have yet to be uncovered[23].

Patterning of the SHF into the transcriptionally distinct AHF and pSHF domains is regulated by retinoic acid signaling-mediated transcription of *Hox* genes, which are

important regulators of anterior-posterior patterning throughout the developing embryo[46]. Retinoic acid is a crucial developmental regulator that functions as a ligand for nuclear receptors, retinoic acid receptors (RAR) and retinoid X receptors (RXR), to control transcription of target genes[47,48]. It is a metabolite of vitamin A, which is first converted to retinaldehyde by alcohol dehydrogenases, followed by irreversible oxidation of retinaldehyde by retinaldehyde dehydrogenases to retinoic acid (RA). The retinaldehyde dehydrogenase family member *Raldh2* is responsible for almost all RA production during early development[48]. At E7.5, *Raldh2* expression occurs in the LPM just posterior to the cardiac crescent; RA generated by *Raldh2* activity in this LPM domain diffuses into the posterior region of the SHF[49] where it regulates transcription of target genes that define the pSHF domain and confer atrial identity[50,51].

In germline deleted *Raldh2*$^{-/-}$ mouse embryos, expression of the *lacZ* transgene driven by a retinoic acid response element *(RARE-LacZ)* and the retinoic acid target gene *Hoxa1* is completely abolished in the pSHF region[50]. Notably, expression of AHF markers including *Fgf8, Fgf10* and *Tbx1* is expanded posteriorly and the posterior expression of the pan-SHF marker*, Isl1* is caudally extended. This expansion is a result of defects in establishing the posterior cell fate, and not due to abnormal posterior migration of AHF cells[50]. These data indicate that RA signaling is important for defining the anterior and posterior limits of the pSHF[50,52]. Phenotypically, *Raldh2*$^{-/-}$ mutants have disrupted atrial and sinus venosus development, consistent with the anteriorized pSHF signature[50,53]. Inhibition of RA-induced transcription during early cardiogenesis leads to defective hearts lacking the atrial chambers, consistent with RA being a driver of atrial cell fate through pSHF cell patterning[51]. However, development of AHF structures is also disrupted upon

perturbing RA signaling; *Raldh2*$^{-/-}$ mutants have OFT defects, likely reflecting the contribution of multiple RA-dependent *Hox* gene-expressing pSHF subdomains to the distal part and inferior wall of the OFT[54–56].

The retinoic acid signaling axis intersects with the T-box transcription factors *Tbx1* and *Tbx5* in establishing the pSHF domain[57]. *Tbx1* regulates the allocation of SHF progenitors to the arterial and venous poles[56] whereas *Tbx5* is a canonical pSHF marker gene critical for venous pole development[16,58] that does not contribute to the arterial pole[13]. RA signaling activates *Tbx5* expression in a subpopulation of *Tbx1*$^{+}$ cells in the posterior SHF region at the heart tube stage; this results in *Tbx5* rapidly downregulating *Tbx1* expression and forming a transcriptional boundary between *Tbx5*$^{+}$ venous pole cells and *Tbx1*$^{+}$ arterial pole cells. Disruption of RA signaling and subsequent inhibition of *Tbx5* activation in this boundary compartment leads to defects in atrial and ventricular septal structures[56]. In summary, RA signaling is critical in early cardiogenesis to appropriately pattern the pSHF and define the progenitor subcompartments that contribute to the arterial and venous pole structures of the heart which are common targets of CHDs.

1.4 Cellular diversity in the developing heart

The developing heart contains a tremendous diversity of cells[6,16]. Among the broad categories of cardiac cell types, such as cardiac progenitor cells (CPCs), myocytes, endothelial cells, smooth muscle cells and mesenchymal cells, each structure of the heart displays region-specific gene expression patterns for these cell types. For example, myocytes of the ventricles, atria, SV, AVC and OFT have distinct gene signatures, reflecting differences in electrophysiology, chamber vs non-chamber myocardium identity

and differential usage of isoforms encoding contractile genes[59]. In addition, these cell subtypes are generated through successive specification events from their respective progenitors via intermediate precursors[60].

During morphogenesis, these heterogeneous cell populations are dispersed throughout the developing heart. Given the exceedingly small size of mouse embryos and the developing cardiogenic region, physical separation of these cell types in order to interrogate their molecular signatures and function is challenging and relies on a priori knowledge of marker genes, which may introduce unintended bias that affects both the questions asked and the conclusions that are drawn. Use of bulk assays to measure cardiac cellular features such as gene expression have not been fruitful, due to the fact that these platforms measure data averaged through the entire population of cells. In light of these limitations, attempts to understand the molecular mechanisms underlying how CPCs are specified to multiple cell subtypes and to comprehensively catalog the extent of cellular heterogeneity of the developing heart, have been hampered. In recent years however, new assays that leverage microfluidic technology to capture and measure molecular features of large numbers of individual cells have gained traction. Of these technologies, the most widely employed and mature modality is single-cell RNA sequencing (scRNA-seq), which has greatly improved our ability to access and study heterogeneity in numerous biological contexts[61].

1.5 Leveraging single-cell transcriptomics to dissect cardiogenesis

Without question, scRNA-seq is revolutionizing our ability to precisely dissect the behaviors, functions and interactions of individual cells in organogenesis, homeostasis and disease pathogenesis at genome scale. This has enabled the discovery of novel cell types, revisions to established lineage hierarchies and identification of novel cell fate determinants when applied to developmental contexts[62–67]. Single-cell transcriptomics is thus accelerating the pace, scale and resolution of biological discovery.

Prior work that applied single-cell transcriptomics to study cardiogenesis between E7.5 and postnatal day (P)21, used plate-based platforms that allow capture of small numbers of cells (1200-2500 cells) with greater coverage and detection of genes[59,68–70]. These studies uncovered novel genes in broad cardiac cell types, described the differentiation and specification defects incurred in cardiomyocytes upon loss of CHD-causing genes and identified heterogeneity and transcriptional features of early committed *Mesp1*$^{+}$ cardiac progenitor cells. A recent study of ~400 single-cell transcriptomes captured from developing human hearts identified the LGR5 gene as a key regulator of progenitors that contribute to the cono-ventricular region[71]. These reports highlighted the value of applying scRNA-seq to quantitatively measure and understand heterogeneity in mouse and human cardiac development transcriptome-wide.

Droplet-based scRNA-seq technologies, where individual cells are co-encapsulated with barcoded beads that capture and encode transcript information, allow acquisition of large numbers of transcriptomes[72]. These technologies can thus further advance our understanding of cardiogenesis, by enabling us to sample cell type heterogeneity more fully, identify rare cell populations and dissect how loss of cardiac

genes differentially impact distinct cell populations. CHDs arise from dysregulation of specific subsets of cell subpopulations. Thus, large-scale single-cell transcriptomics can be leveraged in settings where cardiac development is perturbed and leads to CHDs, to dissect which cell subsets are dysregulated, and precisely determine the functional consequences of that dysregulation. These settings include genetic loss of important core cardiogenic transcription factors. For example, the *Hand2* transcription factor is critical for development and morphogenesis of the SHF, but the mechanisms by which its loss disrupts SHF development are largely unknown. Thus, single-cell RNA sequencing represents a valuable opportunity to determine the consequences of *Hand2* loss in early cardiac development.

1.6 *Hand2* functions in cardiogenesis.

The *Hand2* gene encodes a basic helix-loop-helix (bHLH) transcription factor, that was discovered in a screen to identify bHLH transcription factors expressed during heart development. *Hand2*, then known as *dHand* due to its strong expression in the mouse deciduum, was found to be expressed in the mouse lateral plate mesoderm and cardiac crescent at E7.75, followed by expression throughout the heart tube and SHF at E8.25 and abundant expression in the OFT and neural crest-derived aortic arch arteries, with lower expression in the developing ventricles and atria at E9.25[73,74]. Expression of *Hand2* was also found to be more enriched in the right ventricular chamber, compared to the left ventricle. The expression domain of *Hand2* in these regions spanned both the myocardium and endocardium. Expression of the closely related family member *Hand1,* which has been shown to function redundantly with *Hand2*[73,75,76], was in regions of the

heart that overlapped with *Hand2*, such as the OFT, and was highly enriched in the outer curvature myocardium of the LV[75,77].

The compartment-specific expression pattern of *Hand2* in these early stages of cardiogenesis suggested that it might be important for heart formation. Indeed, this was confirmed upon genetic deletion of *Hand2* in mouse embryos, which resulted in embryonic lethality due to heart failure by E10.5. These *Hand2*-null mutant embryos exhibited a single left-sided ventricular chamber, a shortened OFT and markedly dilated aortic sac. The complete lack of a right ventricle was consistent with the predominant right ventricular expression of *Hand2*. These findings established *Hand2* as a marker and essential regulator of right ventricular chamber formation, while *Hand1* emerged as an essential marker and regulator of left ventricular formation[75,78,79].

Tissue-specific deletion of *Hand2* revealed its critical functions in neural crest cell-derived lineages, the developing limb and the heart[11,80–85]. Notably, *Hand2* deletion in the entire SHF domain phenotypically recapitulated the defect incurred upon its global deletion, revealing its essential role in regulating the SHF[11]. This study implicated *Hand2* as a driver of SHF cell survival, since its loss resulted in increased apoptosis of SHF progenitors in the pharyngeal mesoderm, prior to migration into the arterial pole of the heart. Moreover, temporal deletion of *Hand2* in sub-lineages of the SHF revealed a spectrum of malformations reminiscent of human congenital heart defects.

Over the past twenty years since the discovery of *Hand2*, much has been uncovered about its requirement in multiple tissue lineages, its impact on diverse cellular behaviors in the developing embryo as well as its function as a core cardiac transcription factor[86,87]. However, the transcriptional targets of *Hand2*, as well as a mechanistic

understanding of how it regulates the fate and behavior of distinct cardiac subpopulations have been challenging to uncover, particularly during early stages of cardiogenesis, where progenitor cell gene expression is dynamic and access to individual subpopulations is limited. Thus, one of the major goals of my thesis was to leverage single-cell RNA sequencing to understand how *Hand2* loss impacts early specification and behavior of SHF progenitors in the developing heart.

1.7 Specific aims

The advent of single-cell RNA sequencing technologies that enable the capture of large numbers of transcriptomes represents an invaluable opportunity to study cardiogenesis at high resolution. Accessing the cellular compartments that contribute to specific structures of the heart has thus far been intractable, particular with respect to rare cell populations. Moreover, dissecting how distinct cell types are dysregulated, interrogating the nature of that dysregulation and understanding how it impacts heart formation is now possible, with important implications for understanding the mechanisms underlying manifestation of CHDs in human patients.

In this work, we leveraged droplet-based capture of single-cell transcriptomes to study normal and abnormal cardiac development *in vivo* during murine development. We focused our efforts on the enigmatic early stages of cardiogenesis, when cardiac cell subtypes are undergoing active specification from their respective progenitor compartments and the developing heart is undergoing major morphogenetic changes. Our goal was to categorize all the cell types and cell states that we captured in normal development and identify fate-determinants that drive specification of multiple cardiac

subtypes from the same progenitor domain. We then used this high-resolution molecular framework to dissect how loss of the lineage-specifying transcription factor, *Hand2*, disrupts these cell populations, resulting in a complex developmental defect. Collectively, this work presents a valuable resource for the cardiac development field, uncovers how loss of a broadly expressed transcription factor leads to differential perturbation in distinct cell types and demonstrates how single-cell RNA sequencing can be leveraged to reveal the molecular mechanisms that are dysregulated in congenital heart defects.

Chapter 2

Analysis of wild-type cardiogenesis

2.1 Rationale

The heart develops from diverse cell lineages that are specified from two pools of cardiac progenitor cells (CPCs), the FHF and SHF, and multipotent neural crest cells, yet our understanding of the molecular regulators that specify distinct fates from these domains is limited. While several decades of research in the cardiac development field have uncovered lineage-specific markers and important cell-fate regulators of these populations, an unbiased transcriptome-wide characterization of the cellular diversity of the heart has not been possible until now, due to limitations of bulk technologies. The first goal of my thesis was to use single-cell RNA sequencing to capture cardiac cell transcriptomes from wild-type (WT) mouse embryos, in order to comprehensively describe and interrogate the cellular heterogeneity of the developing heart.

2.2 Mapping early cardiogenesis at single-cell resolution

To identify transcriptional features of cardiac cell fate specification and morphogenesis at the single-cell level, we used the 10X genomics Chromium platform[72] to capture transcriptomes of cells from the cardiogenic region in mouse embryos at three developmental stages: 1) as CPCs begin to differentiate and form a late cardiac crescent at E7.75; 2) as the FHF forms a linear, beating heart tube and the SHF migrates into the anterior and posterior poles of the tube at E8.25; and 3) as the heart tube loops and incorporates the SHF-derived right ventricle and atrial cells with the FHF-derived left ventricle and atrial cells at E9.25 (**Fig. 2.1a, b**).

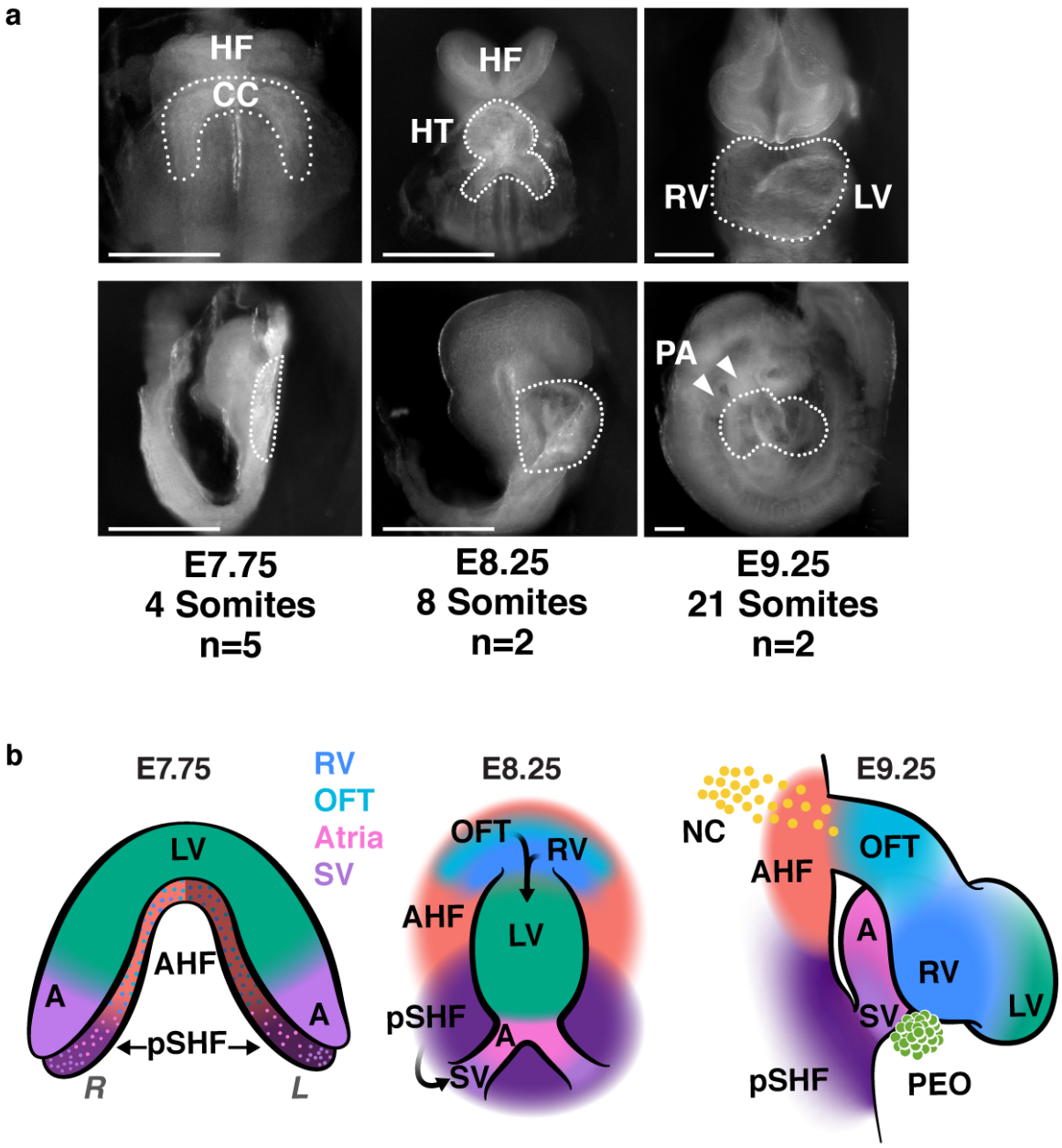

Figure 2.1: Capture of cardiac populations in early mouse embryos.
a, Representative images of mouse embryos at E7.75, E8.25 and E9.25 used for cell collection, with micro-dissected regions indicated, in frontal view (top) and right sagittal view (bottom). Scale bar, 200 μm. **b**, Spatial organization of captured cardiac cell populations at each stage. Darker shaded region on left side of SHF at E7.75 indicates left-right asymmetric patterning. Dots in the SHF at E7.75 represent differentiating RV, OFT, atrial and SV cells as indicated by the color labels. HF, head folds; CC, cardiac crescent; HT heart tube; RV, right ventricle; LV, left ventricle; PA, pharyngeal arches; FHF, first heart field; SHF, second heart field; AHF, anterior heart field; pSHF, posterior second heart field; NC, neural crest cells; OFT, outflow tract; SV, sinus venosus; A, Atria; PEO, proepicardial organ containing epicardial cells.

We sequenced the transcriptomes of 36,777 cells, including ectoderm and endoderm progenitors that were unintentionally captured during dissection of the mouse cardiogenic region, over the three developmental time-points (**Fig. 2.2a, b**). We computationally excluded these ectoderm and endoderm cells as well as contaminating blood cells that were captured at E9.25 (**Fig. 2.2c**). We applied a graph-based clustering approach on the remaining 21,366 neural crest-derived mesenchyme cells and mesoderm populations. This analysis partitioned the cells into 7 broadly defined populations: multipotent *Isl1*$^{+}$ progenitors, endothelial/endocardial cells (*Cdh5*$^{+}$*/Emcn*$^{+}$*/Plvap*$^{+}$*)*[88–90]*,* epicardial cells (*Tbx18*$^{+}$*/Wt1*$^{+}$*)*[91,92]*,* myocardium (*Myh7*$^{+}$*/Myh7*$^{+}$*/Tnnt2*$^{+}$*),* neural crest-derived mesenchyme cells (*Dlx2*$^{+}$*/Dlx5*$^{+}$)[93,94], paraxial mesoderm (*Tcf15*$^{+}$*/Meox1*$^{+}$)[95,96] and posterior lateral plate mesoderm (*Hoxb6*$^{+}$*/Foxf1*$^{+}$)[97,98] (**Fig. 2.2d, e**). We subdivided the multipotent *Isl1*$^{+}$ progenitors, endothelial/endocardial cells and myocardium populations to appreciate additional heterogeneity (**Fig. 2.3**). To assign identities to these subpopulations, we cross-referenced the most highly and uniquely expressed genes in each subpopulation with known cardiac subtype markers and *in situ* hybridization data from the literature (**Supplementary Table 2.1**). We further validated several of these marker genes by fluorescence *in situ* hybridization, which provided spatial resolution (**Fig. 2.4).**

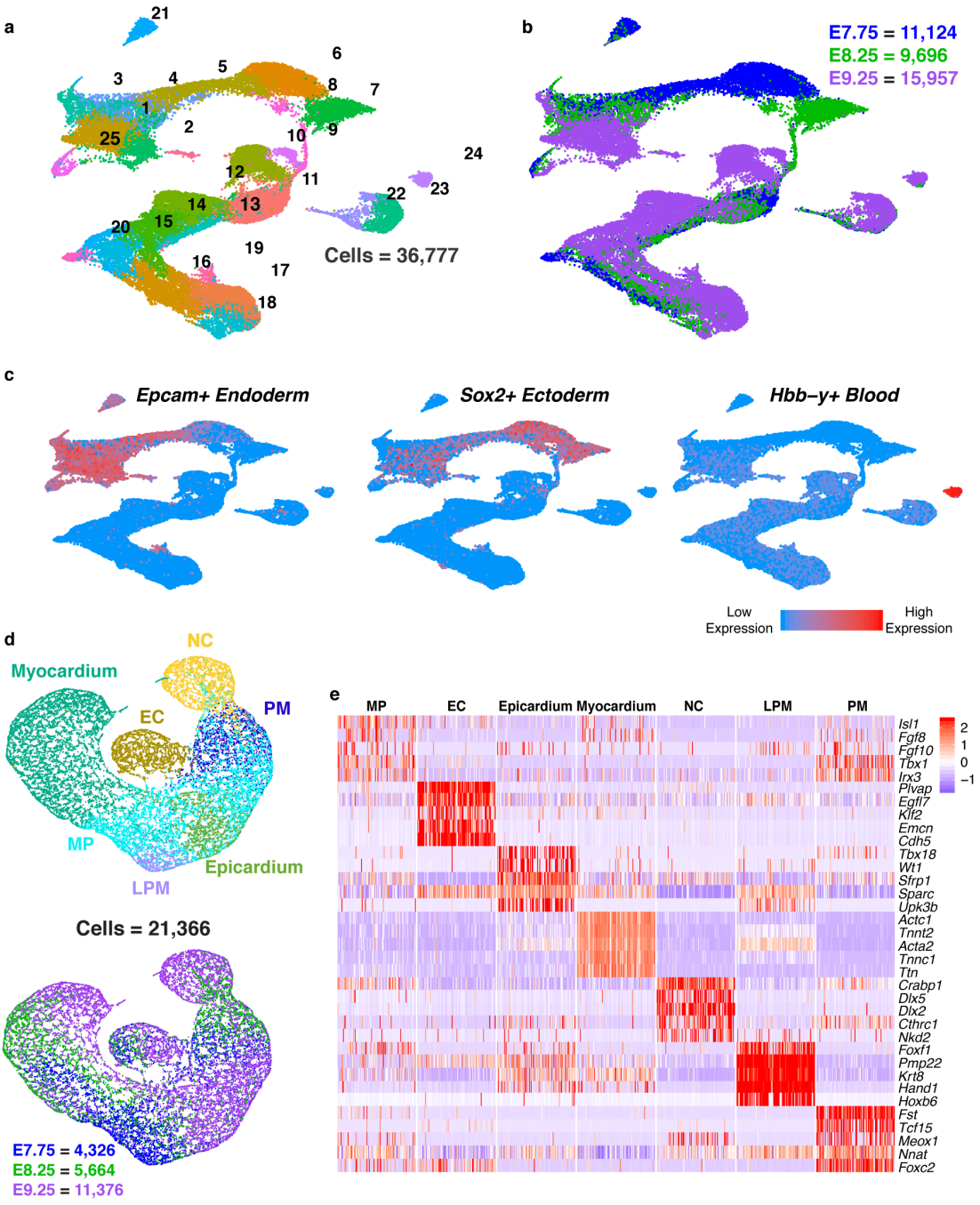

Figure 2.2: Single-cell RNA-seq reveals heterogeneity of cardiac populations. **a,** UMAP plot of all captured cell populations colored by cluster and **b,** stage of collection. **c,** UMAP plot showing expression of markers used to identify contaminating populations. **d**, UMAP plot of mesodermal and neural crest populations colored by cluster identity and embryonic stage of collection. **e**, Expression heatmap of marker genes of populations. Statistics for differential gene expression tests were applied to n = 21, 366 cells. Data are shown for 100 cells subsampled from each population. Scale indicates Z-scored expression values. MP, multipotent progenitors; EC, endocardial or endothelial cells; PM, paraxial mesoderm; LPM, lateral plate mesoderm.

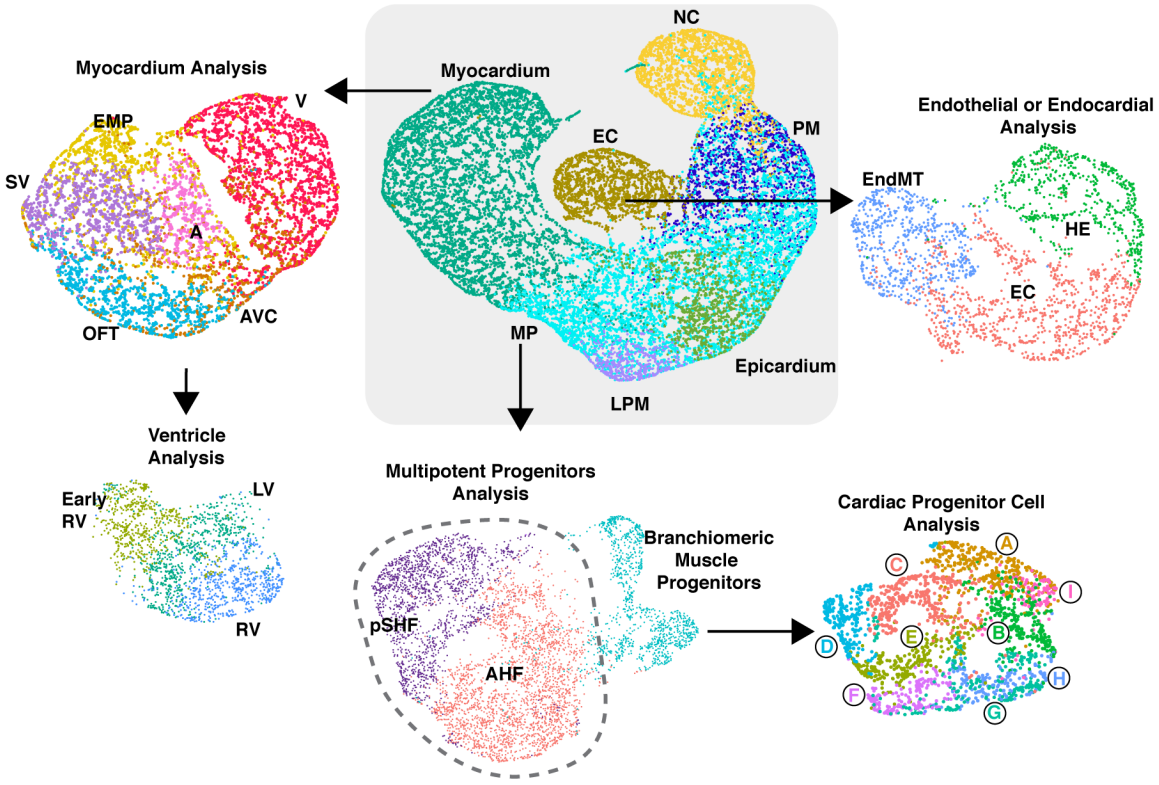

Figure 2.3: Focused analyses of cardiac populations.
Schema of progressive subdivisions of broadly clustered cell populations from Fig 2.2d.

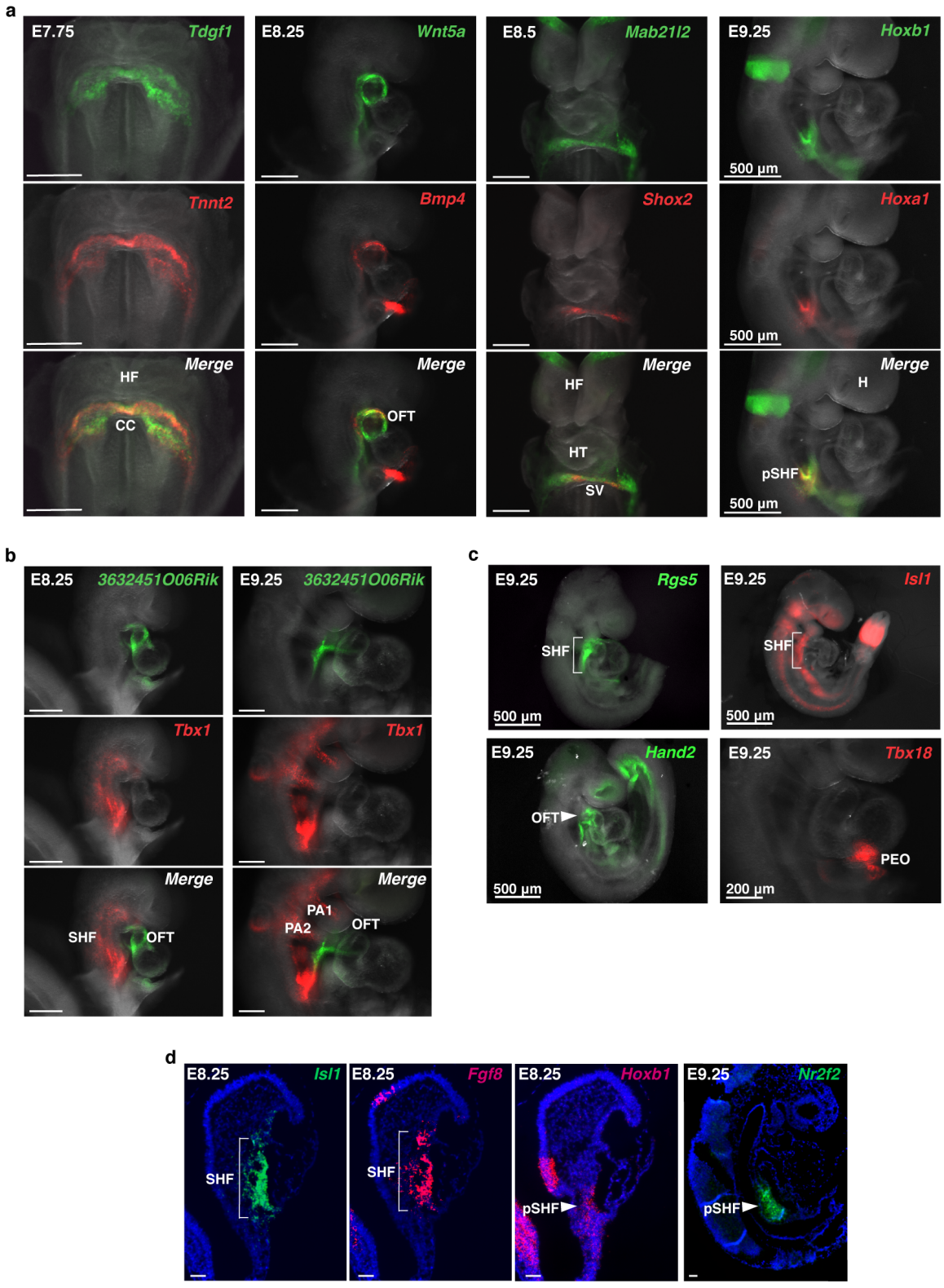

Figure 2.4 legend continued on next page

Figure 2.4: Spatial validation of marker gene expression by *in situ* hybridization. **a,** Ventral view of *Tdgf1* and *Tnnt2* expression in the CC, right lateral views of *Wnt5a* and *Bmp4* in the OFT, *Mab21l2* and *Shox2* in the SV*,* and *Hoxa1 and Hoxb1* in the posterior pSHF by *in situ* hybridization at days indicated that informed assignment of population identities**. b,** Expression of *Tbx1* in the SHF and pharyngeal arches PA and novel unannotated gene *3632451O06Rik* in the OFT at E8.25 and E9.25. **c,** Expression of *Rgs5* and *Isl1* in the SHF*, Hand2* in the SHF and OFT and *Tbx18* in the PEO of E9.25 embryos. Scale bars indicate 200 μm unless otherwise noted. **d,** *In situ* hybridization of mRNA expression of *Isl1, Fgf8* and *Hoxb1* at E8.25 and *Nr2f2* at E9.25 in right lateral histologic sections. n=2 independent embryos per gene for all panels. Scale bars indicate 50 μm. H, head.

Within each broad population of cells, further sub-clustering revealed distinct cell types characteristic of unique progenitor pools that are spatially organized (**Supplementary Table 2.1**). Among the endocardial/endothelial lineage, we captured three subpopulations of the endocardial/endothelial lineage: hematoendothelial progenitors (*Etv2*^{high}/ *Tal1*^{high}/ *Cdh5*^{low})[99], specified endothelial/endocardial cells (*Cdh5*^{high}/*Pecam1*$^{+}$/*Emcn*^{high}), and endocardial cells initiating an endothelial-to-mesenchymal transition program typical of valve development (*Hand2*^{high}/*Msx1*^{high}/*Gata4*^{high})[22] (**Fig. 2.5 a, b**). The *Isl1*$^{+}$ multipotent progenitor population subdivided into the AHF, pSHF, and branchiomeric muscle progenitors that reportedly share a common origin with SHF cells (**Fig. 2.5c, d**)[100,101].

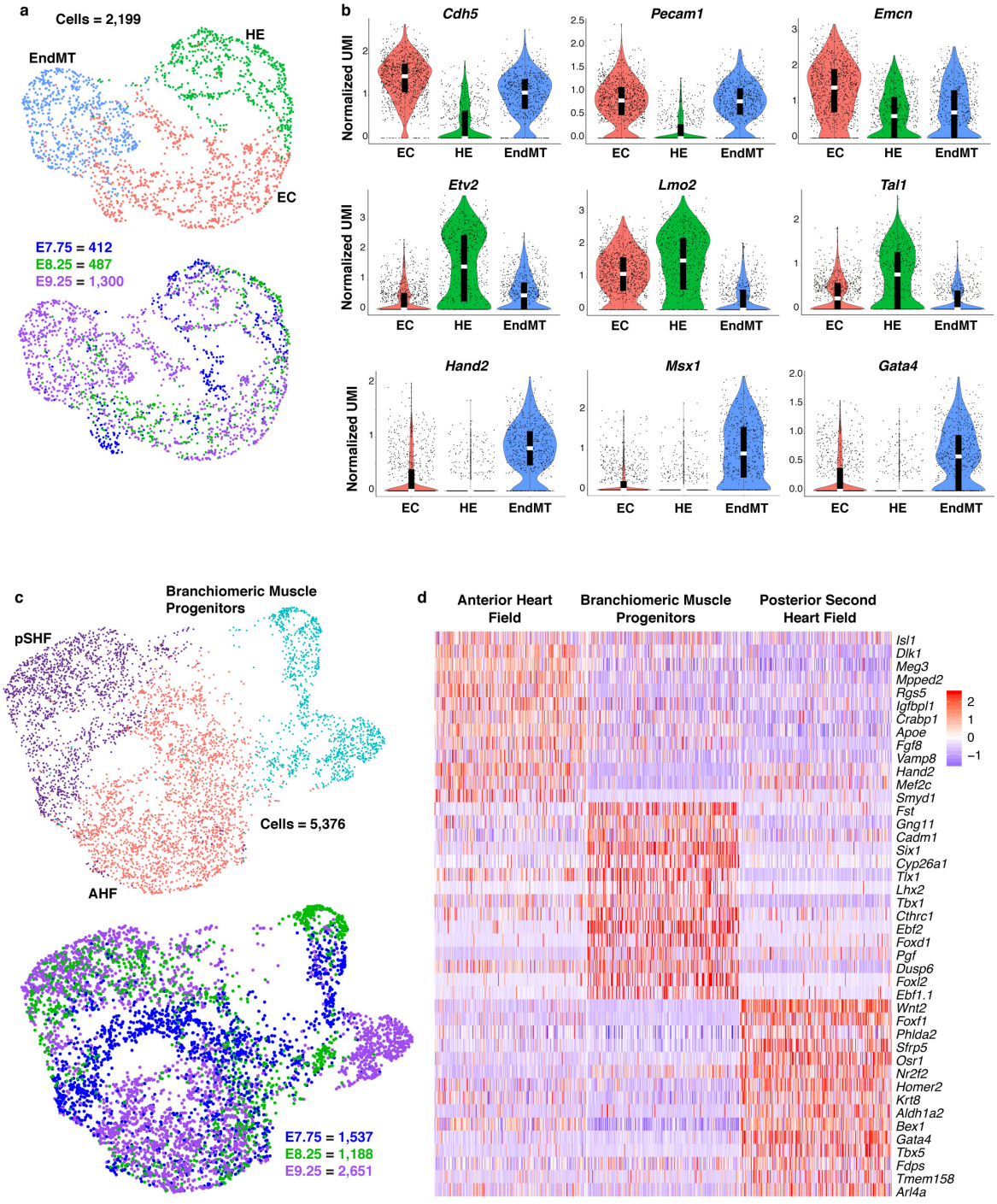

Figure 2.5: Heterogeneity in endocardial or endothelial cells and multipotent progenitor populations.

a, UMAP plot of reclustered endocardial or endothelial population colored by cluster and embryonic stage of collection. **b,** Violin plot of markers indicating distinct subpopulations of endocardial or endothelial cells. Summary statistics reported in all violin plots: the

center white line represents median gene expression and the central black rectangle spans the first quartile to the third quartile of the data distribution. The whiskers above or below the box indicate value at 1.5x interquartile range above the third quartile or below the first quartile. Statistics for differential gene expression tests were applied to n=2,199 cells. **c,** UMAP plot of reclustered multipotent progenitor populations colored by cluster and embryonic stage of collection. **d,** Heatmap showing curated list of marker genes that identify pSHF, AHF and branchiomeric muscle progenitors. Scale indicates Z-scored expression values. Statistics for differential gene expression tests were applied to n=5,376 cells. HE, hemato- endothelial progenitors; EC, endocardial or endothelial cells; EndMT, endothelial-mesenchymal transition cells; AHF, anterior heart field; pSHF, posterior second heart field.

We also identified transcriptomes representative of ventricular (*Irx4*$^{+}$)[102,103], atrial (*Cav1*$^{+}$/*Kcna5*$^{+}$)[104,105], sinus venosus (*Shox2*$^{+}$)[106], atrioventricular canal (AVC, *Rspo3*$^{+}$/*Bmp2*$^{+}$)[107,108] and OFT *Tdgf1*^{high}/*Rspo3*$^{+}$/*Irx4*$^{-}$)[107,109] myocardium (**Fig. 2.6a, b**). We defined the gene signatures of each of these sub-populations as these transcriptomes provide the first insight into the broader genes expressed in specific cell types of the heart. For example, a novel, unannotated single-pass transmembrane domain protein, *3632451O06Rik*, emerged as highly enriched in OFT myocardium (**Fig. 2.6b**). Although the human ortholog of *3632451O06Rik* has been shown to interact with *mTORC2*[110], the putative cardiac function of this gene and its role in OFT cells has not been interrogated. We also found highly enriched expression of the Regulator of G protein signaling gene, *Rgs5,* in the second heart field and OFT (**Fig. 2.4c**). Signaling mediated by G-proteins is critical for development of the heart, particularly for SHF-derived structures[111–114], and *Rgs5* may be involved in regulating these pathways. This dataset therefore represents a catalog of cardiac cell states that arise during embryonic development, along with their transcriptomes, providing a foundation to study the transcriptional dynamics underlying specification of cardiac subtypes, as well as to uncover novel cell type specific genes.

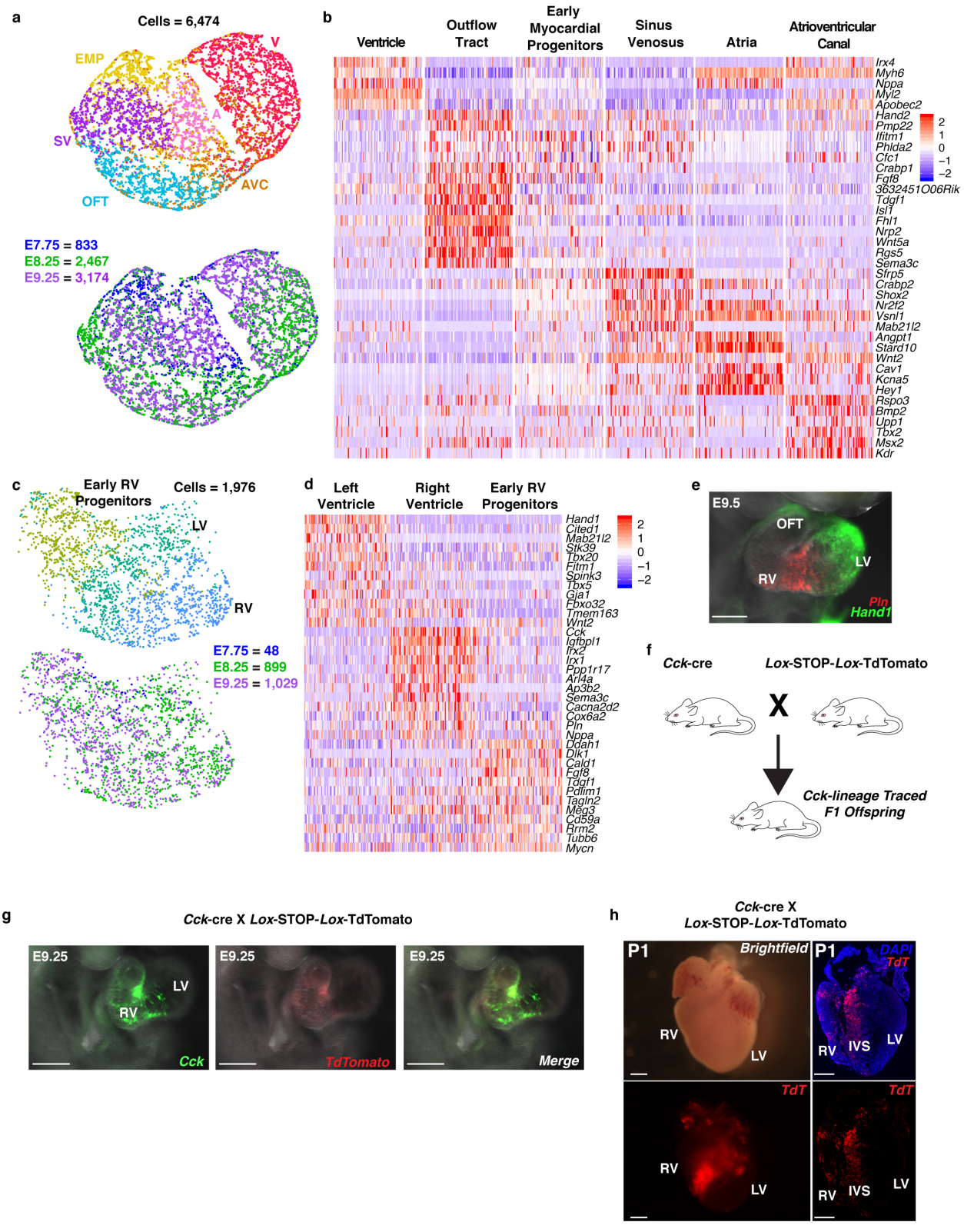

Figure 2.6 legend continued on next page

Figure 2.6: Focused analyses of myocardial populations and spatial validation of right ventricle markers.

a, UMAP plot of reclustered "Myocardium" population colored by cluster and embryonic stage of collection. **b,** Heatmap of highly and uniquely expressed genes in myocardial subpopulations. Scale indicates Z-scored expression values. Statistics for differential gene expression tests were applied to n=6,474 cells. **c,** UMAP plot of reclustered ventricle populations colored by cluster and embryonic stage of collection. **d,** Heatmap showing curated list of genes that identify left ventricle (LV), right ventricle (RV) and early RV progenitors. Scale indicates Z-scored expression values. Statistics for differential gene expression tests were applied to n=1,976 cells. **e,** mRNA expression of left ventricle marker *Hand1* (green) and *Pln* (red) in frontal view of the E9.5 heart showing enrichment in right ventricle region by whole mount *in situ* hybridization. n=2 independent embryos per gene, Scale bar, 200 μm. **f,** Breeding scheme for lineage-tracing *Cck* expressing cells. **g**, mRNA expression of endogenous *Cck* and *TdTomato* driven by *Cck*-cre transgene at E9.25 in right oblique view of the heart. n=2 independent embryos per gene; scale bar, 200 μm. **h,** Expression of *TdTomato* in whole-mount and sectioned postnatal day 1 (P1) heart from *Ai14xCck*-cre lineage-traced mice showing location of progeny of *Cck*-expressing cells. Left panels show brightfield view (top) or TdTomato (bottom) of whole-mount P1 heart; right panels show sections of TdTomato and DAPI (top) or TdTomato alone (bottom) in postnatal day 1 (P1) heart section. n=2 independent embryos. Scale bar, 100 μm. EMP, early myocardial progenitors; AVC, atrioventricular canal; IVS, interventricular septum.

While the left and right ventricles perform similar functions, they arise from distinct progenitors. Moreover, dysregulation of genes that are uniquely enriched in either of these populations leads to chamber-specific congenital malformations[74,115]. In order to identify novel genes characteristic of the LV and RV, we subdivided the ventricle (V) population (**Fig. 2.6a**) comprising 1,976 cells to three populations, one of which represented the LV based on its highly enriched expression of the genes *Hand1* and *Cited1* **(Fig. 2.6c, d)**[116]**.** The remaining two *Hand1*$^{-}$/*Cited1*$^{-}$ populations are RV, where one represents a less differentiated population (termed Early RV progenitors) due to its lower expression of ventricular myocardium genes such as *Myl2, Tnni3* and *Nppa* and the higher expression of AHF progenitor genes such as *Fgf8*[117] (**Fig. 3.6d**). Several genes were enriched in only one chamber, consistent with their distinct origins and physiology

(**Supplementary Table 2.1**). Surprisingly, Phospholamban (*Pln*), a critical regulator of calcium handling[118], and the Cholecystokinin gene (*Cck*), encoding an intestinal and neuronal peptide hormone[119], were predominantly expressed in RV cells (**Fig. 2.6d, e**). *In situ* hybridization confirmed the RV dominance of *Cck* and also revealed enrichment in the region of the future ventricular septum at E9.25 (**Fig. 2.6f, g**). Lineage tracing of *Cck* expressing cells and their progeny using a constitutive *Cck*-IRES-*Cre* transgenic mouse crossed with a floxed TdTomato reporter mouse further established their enrichment in the RV, particularly the trabecular myocardium, and the IVS, with some expression in parts of the atria (**Fig. 2.6g, h**). These data demonstrate the power of leveraging single-cell transcriptomics to reveal unique genes that characterize diversity among highly similar cell subtypes with distinct physiologies.

A myocardium population that we could not ascribe to a known cell type comprised a large proportion of E7.75 cells and expressed genes of multiple myocardium subtypes such as *Irx4* (ventricle) and *Angpt* (atrial)[120] (**Fig. 2.6a, b**). We hypothesized that these cells represent a mixed population of early myocardium progenitors (EMP) including FHF cells. To test this hypothesis, we used pseudotemporal ordering, which is a computational measure of the progress a cell makes along a differentiation trajectory[121]. We constructed a trajectory of all the myocardium populations described above, which revealed 6 major cell states (**Fig. 2.7a-e**). The majority of cells from the EMP population were placed at the start of the trajectory along with a subset of cells classified as SV, suggesting that the two populations were in a transcriptionally similar cell state, while the remaining SV cells were placed with atrial and AVC cells along the trajectory, consistent with their overlapping gene expression signatures (**Fig. 2.6b**). Notably, EMP cells from the later time points

were represented in all other myocardial cell states, consistent with our hypothesis (**Fig. 2.7b).** Additionally, the early RV progenitor cells identified in the ventricle analysis were placed with the OFT state (State 4), a state comprising differentiated LV/RV cells (State 6) and an intermediate state (State 5) that expressed lower levels of differentiated ventricle genes, such as *Nppa*[122] and higher levels of *Fgf8*, supporting their broad designation as early RV progenitors (**Fig. 2.7d, f**). We used the pseudotime analysis to identify several genes that are dynamically expressed and enriched in the EMP/SV and OFT states such as *Abra, Dlk1, Cfc1, Ifitm1, Mtus2, Marcks* and *Phlda2* (**Fig. 2.7g**). This analysis also identified the *Hand2* gene, which is typically considered a marker of ventricular identity[77] to be more highly expressed in outflow tract myocardium cells than ventricle cells. *In situ* expression analysis at E9.25 also confirmed the enriched expression of *Hand2* in the OFT (**Fig. 2.4c**).

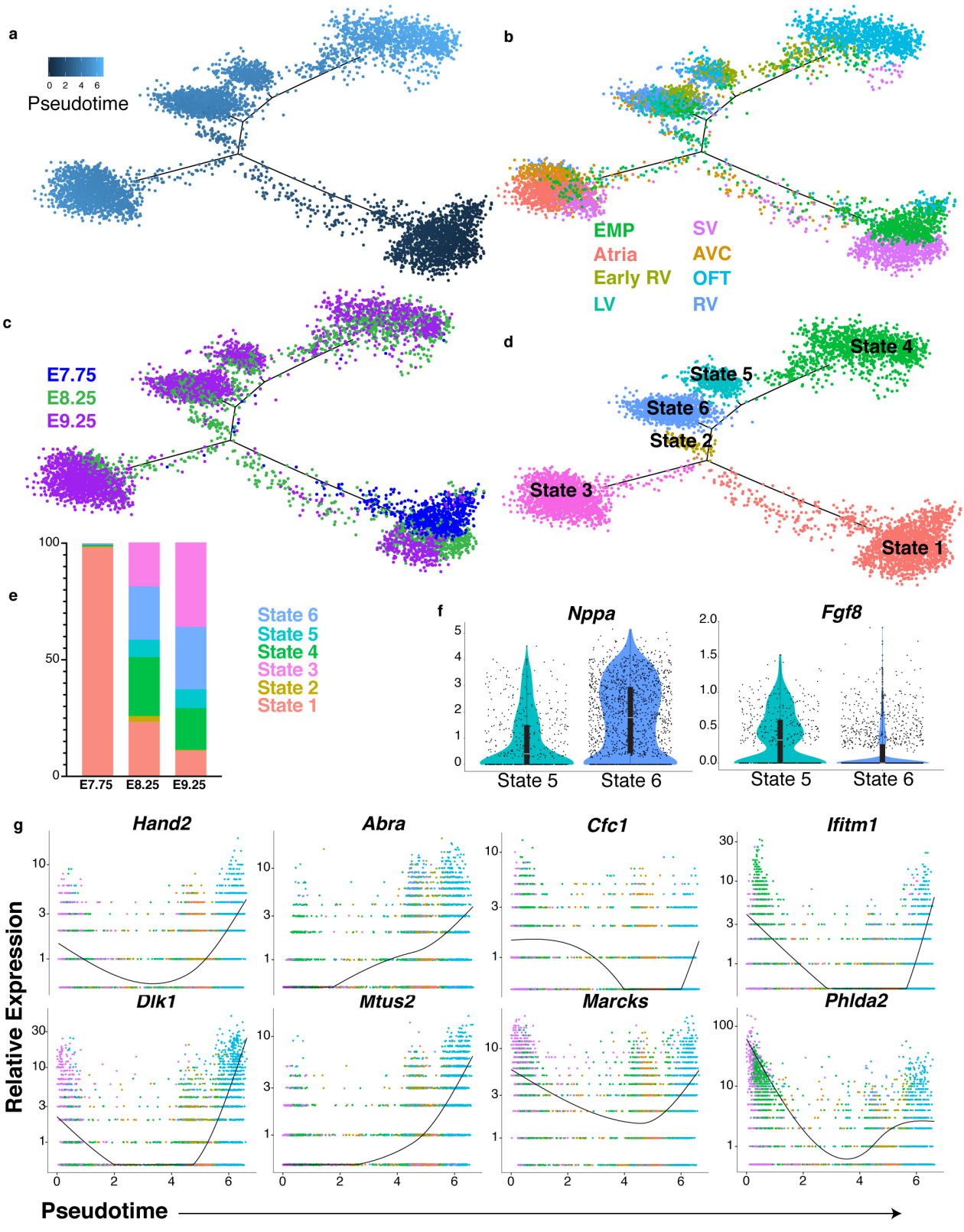

Fig. 2.7 legend continued on next page

Figure 2.7: Pseudotemporal ordering of myocardium populations.
Pseudotime trajectory of myocardium populations colored by **a,** pseudotime value, **b,** cluster identity, **c,** embryonic stage of collection and **d,** cell state. Pseudotime trajectory analysis was applied to n=6,474 cells. **e,** Percentage of cells in each state that were captured at E7.75, E8.25 or E9.25. **f,** Violin plots showing expression of *Nppa* and *Fgf8* in State 5 and State 6 from pseudotime trajectory in (**d**). **g,** Expression dynamics of select genes along the myocardium pseudotime trajectory. Dots represent individual cells at indicated relative expression levels colored by cluster identity. Statistics for differential gene expression tests were applied to n = 455 cells from each state. Bonferroni correction adjusted p-value < $1x10^{-4}$ (Wilcoxon rank sum test, two-sided).

2.3 A framework to investigate gene function in distinct cell types

CHDs are caused by disruption of relatively small subpopulations of cardiac cells during development and substantial efforts are ongoing to identify genetic variants that underlie the large majority of CHD cases[5]. Many novel genes previously not implicated in heart development have been identified from these studies, but the cell types in which they function are unknown. A high-resolution molecular framework of cardiogenesis such as the one presented here can be leveraged to link the expression of such novel genes to specific cell types. To that end, we cross referenced the marker genes identified in all captured populations in our dataset against genes identified in CHD patients with damaging de novo mutations as assessed by the MetaSVM algorithm[5] and identified several that demonstrated cell type specific or enriched expression patterns. These included *Upp1* and *Flt4* in the endocardium/endothelium and *Prkaa2, Rrad* and *Ank3* in the myocardium (**Fig. 2.8**). Thus, our single cell dataset represents a valuable resource to identify cell populations that may be dysregulated and contribute to CHD pathogenesis.

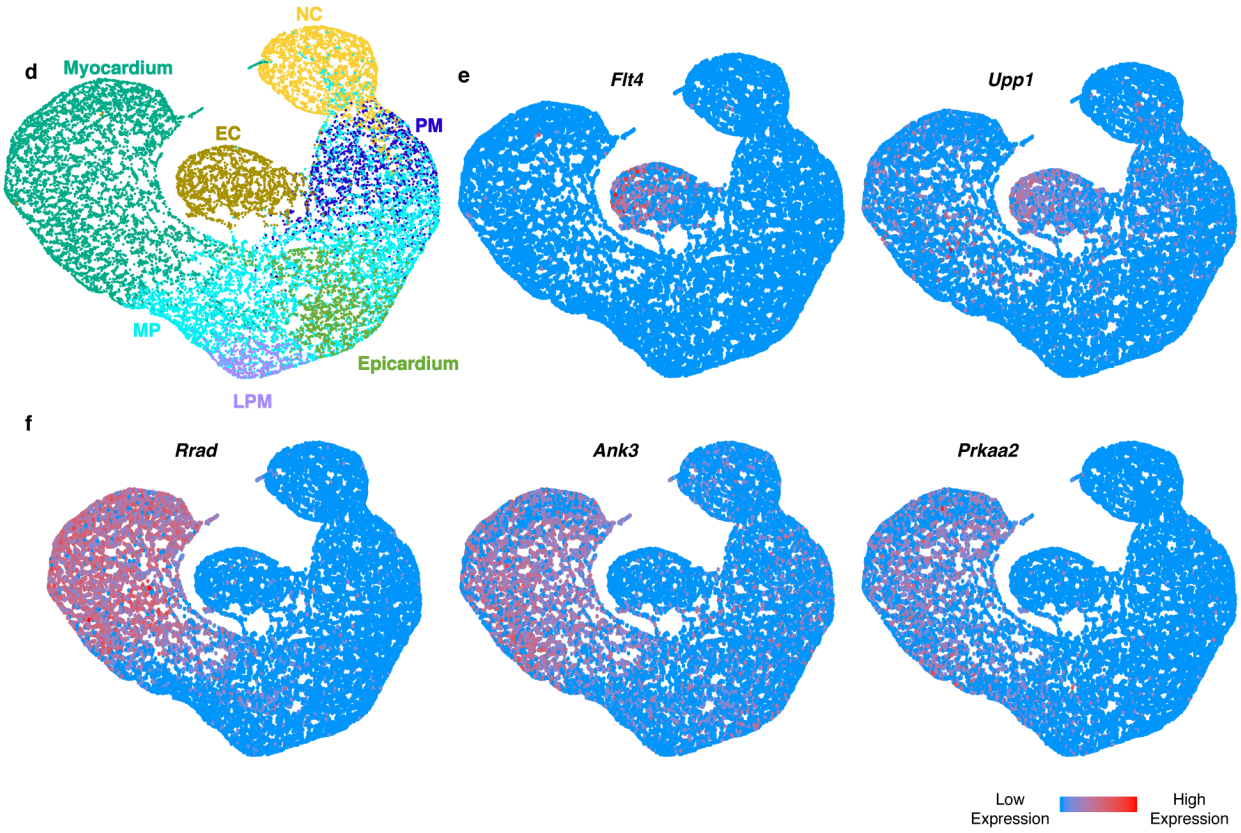

Figure 2.8: Novel genes associated with CHD are enriched in specific cardiac populations.
UMAP plot of all mesodermal and neural crest populations captured at E7.75, E8.25 and E9.25 colored by cluster identity and showing expression of *Flt4* and *Upp1* in endocardial or endothelial population and *Rrad, Ank3* and *Prkaa2* in sub-populations of the Myocardium.

2.4 Dissecting additional heterogeneity of cardiac progenitor cells

Previous studies employing clonal genetic fate mapping and scRNAseq of the earliest *Mesp1*$^{+}$ cells demonstrated that these cells are specified towards committed CPCs fated for distinct anatomic regions and lineages between E6.25 and E7.5[13,14,69]. Given that our dataset contained large numbers of CPCs, compared to previous single-cell CPC analyses[59,68,69], we sought to characterize the additional heterogeneity within this compartment. We focused on the AHF and pSHF populations captured at E7.75 and E8.25 (**Fig. 2.5c**), and this analysis further subdivided the CPCs into nine populations

(**Fig. 2.9a, Supplementary Table 2.1**). Populations A–C were of pSHF origin, whereas E–I were of AHF origin. The pSHF cluster A and AHF cluster I were derived from E7.75 and >50% of cells in each population co-expressed the left sided genes *Nodal, Lefty2* and *Pitx2*,[37] (**Fig. 2.9a, b)** suggesting that these populations contained left-right asymmetrically patterned progenitor cells. A differential gene expression test between *Pitx2*-positive (normalized UMI > 0.1) and *Pitx2*-negative (normalized UMI < 0.1) cells in A and I identified genes with putative asymmetric expression associated with these left-right asymmetric markers (**Supplementary Table 2.1).** Population D, largely from E7.75, had enriched expression of LV genes such as *Hand1, Mab21l2* and *Irx4*[75,103,123]*,* along with sarcomeric genes, and minimal levels of *Isl1*, which indicates these cells are early FHF progenitors (**Fig. 2.9b**). Several populations of the AHF and pSHF lineages expressed genes indicative of myocardial differentiation (*Tnnt2, Actc1, Myh7*) prompting us to perform pseudotime analysis to gain insights into the distinct differentiation dynamics of the AHF and pSHF (**Fig. 2.9c, d, e**). The resulting trajectory suggests that pSHF clusters A and B might represent sequential stages of a differentiating population, while pSHF cluster C encompasses a continuum of cells that are differentiating to myocardium (**Fig. 2.9b, d**). Moreover, the AHF population F appears to represent the earliest outflow tract cells due to its expression of genes such as *Hand1*[77], *Bmp4* (**Fig. 2.4a**) and *Dlk1* (**Fig. 2.7g**). This analysis identified the gene *Upp1*, which was previously unknown to function in the heart, enriched in differentiating pSHF myocardium. Additionally, the retinoic acid (RA) binding protein *Crabp1*, which inhibits RA signaling, was highly enriched in the differentiating AHF cells (**Fig. 2.9b**)[124].

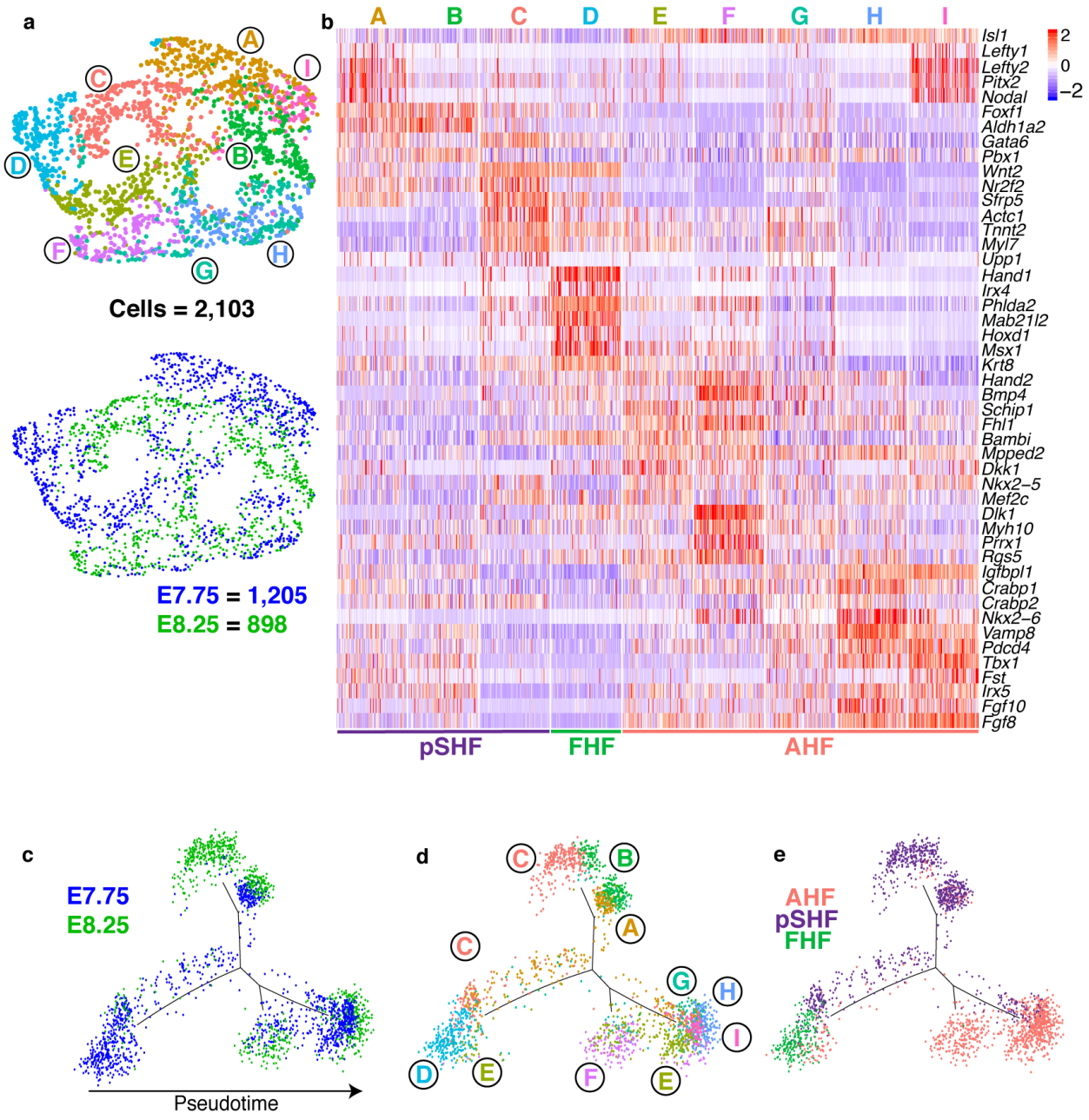

Figure 2.9: Analysis of cardiac progenitor cell populations reveals early specification dynamics of myocardial subtypes.
a, UMAP plot of AHF and pSHF subclusters colored by cluster identity and embryonic stage of collection. **b,** Expression of indicated highly and uniquely expressed genes in subpopulations from (**a**). Scale indicates Z-scored expression values. Pseudotime trajectory of CPCs colored by **c,** embryonic stage of collection, **d,** population identity and **e,** AHF, pSHF or FHF origin.

2.5 Diversity of endodermal populations associated with cardiogenic mesoderm

As CPCs form the cardiac crescent, induction of cardiac gene expression is driven by FGF and BMP signals secreted from the adjacent anterior endoderm[125]. We hypothesized that there may be other endodermal secreted factors expressed by the FGF- and BMP- secreting cells. To test this, we analyzed 915 *Sox17*$^{+}$*/Foxa2*$^{+}$ endoderm cells captured at E7.75 and identified five subpopulations (**Fig. 2.10a, Supplementary Table 2.1**. Expression of cardiac inducers *Bmp2*, *Bmp4* or *Fgf8* was enriched in specific subpopulations, suggesting that cardiac crescent cells receive combinatorial inductive signals from distinct endodermal cell subpopulations (**Fig. 2.10b**). Cluster 5 was highly enriched in genes characteristic of mesendoderm (*Bry, Nog*[126] *and Shh*[127]) while Cluster 1 appears to represent the cardiogenic anterior visceral endoderm, expressing genes such as the Wnt antagonist *Dkk1* and factors such as *Otx2* and *Gsc*[128] (**Fig. 2.10c**), revealing the previously undescribed transcriptomes of this, and other, specific endodermal populations. For example, we identified *Wnt5a,* a non-canonical *Wnt* ligand, enriched in multiple clusters that co-expressed *Bmp4, Bmp2* and *Fgf8 (***Fig. 2.10b)**.* While *Wnt5a* is expressed and required in the SHF[129], its expression in the endoderm at E7.75 was unappreciated, as broader transcriptomes of cardiogenic endoderm were not formerly accessible.

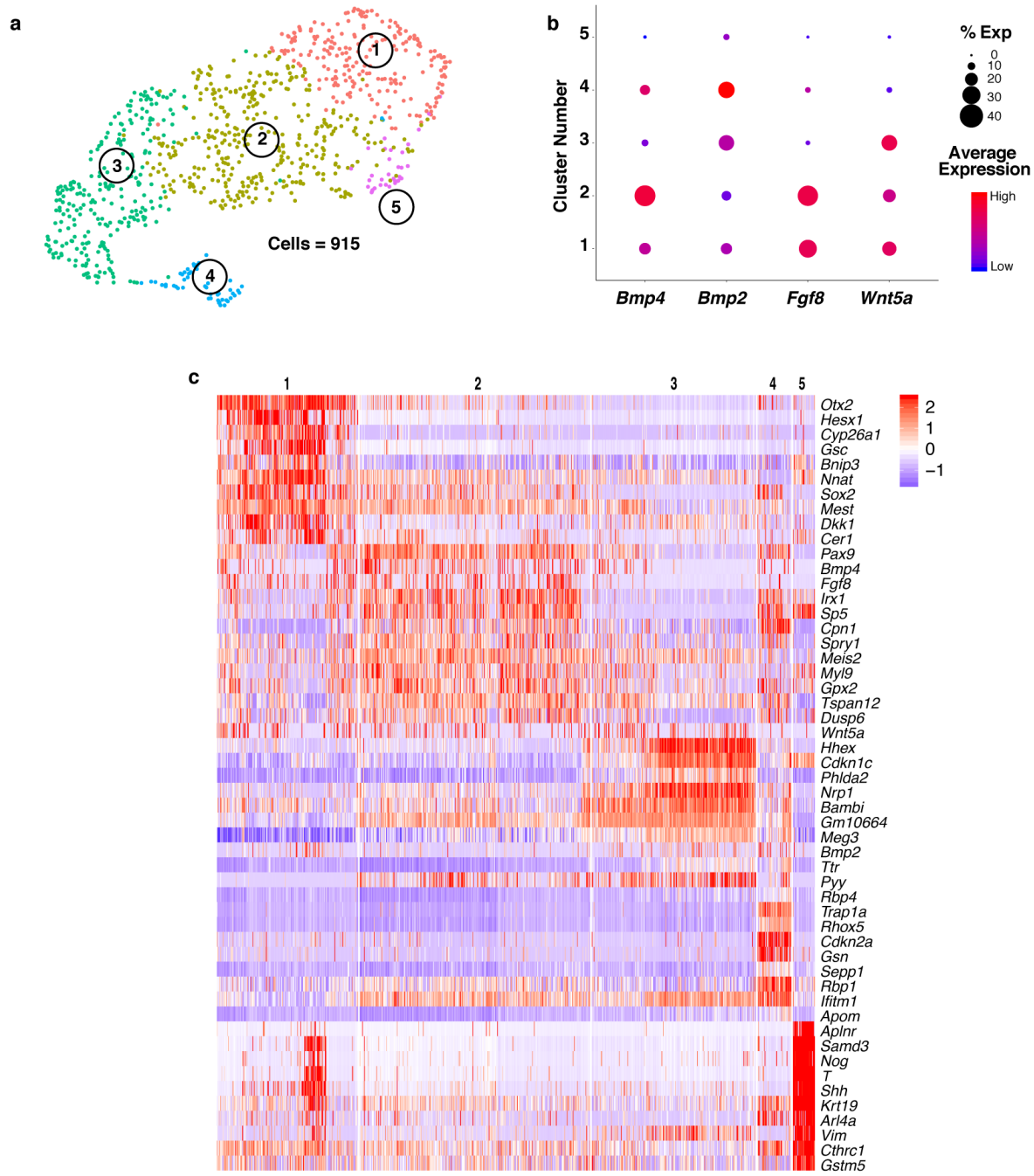

Figure 2.10: Endoderm populations adjacent to the cardiac crescent.

a, UMAP plot of endoderm populations captured at E7.75. **b**, DotPlot of expression patterns of known and novel endodermal secreted factors. The size of the dot indicates the percentage of cells expressing that gene within a cluster, while the color encodes the average expression level of that gene within a cluster. **c**, Expression heatmap of the top ten marker genes of each endodermal population. Scale indicates Z-scored expression values. Statistics for differential gene expression tests were applied to n = 915 cells.

2.6 Predicting cell fate determinants of AHF derivatives

The AHF gives rise to chamber (RV) and non-chamber (OFT) myocardial lineages. However, the molecular regulators governing differential specification of these lineages are unknown. To identify such lineage specifiers using our single-cell RNA sequencing data, we applied a Boolean network-based computational method that takes as input a list of differentially expressed genes between a parent progenitor cell population and the two daughter cell populations representing different fates to systematically predict cell-fate determinants[130,131] (see Methods and Supplementary Table 2.1). The procedure was applied to E7.75 and E8.25 AHF, RV and OFT cells and identified *Irx4* and *Plagl1* as lineage specifiers of the RV, while *Hand2, Tead2* and *Arid3b* were determined as specifiers of the OFT. *Irx4* is an established specifier of ventricular identity[102,103], thus its identification as a cell fate determinant for RV cells confirms the validity of this analysis. *Arid3b* is important for the deployment of SHF progenitors[132] and its deletion results in OFT shortening, supporting its identification as a determinant of OFT fate. The prediction of *Hand2* as a specifier of OFT cells from our lineage-determinant analysis is consistent with its differential, enriched expression in the OFT myocardium population (**Fig. 2.5b**). However, it was surprising that *Hand2* was predicted as a lineage specifier for the OFT, but not RV, myocardium as its global deletion causes lethality by E10.5 secondary to severe RV hypoplasia[74]. This phenotype is recapitulated upon conditional deletion of *Hand2* in the SHF, underscoring the requirement of *Hand2* in this progenitor compartment[11]. How RV hypoplasia occurs in this genetic model is unknown, in part due to a previous inability to access individual progenitor cells that may be disrupted during

specification. We address this gap in knowledge using single-cell RNA sequencing in Chapter 3.

2.7 Conclusions and Discussion

In this work, we compiled a high-resolution map of the cellular heterogeneity in early cardiogenesis. We identified novel cell-type enriched genes, described previously unrecognized heterogeneity in cardiac progenitor compartments, and uncovered lineage-specifying regulators of the RV and OFT myocardium from the AHF. This endeavor represents the largest comprehensive catalog of cell subtypes in the developing heart and will be a valuable resource to the cardiac development community. Accordingly, we have partnered with Dr. Maximilian Haeussler and Dr. Matthew Speir to upload and format all processed datasets described herein on the UCSC cell browser at https://mouse-cardiac.cells.ucsc.edu. The browser is a novice-user-friendly platform for wide dissemination and interrogation of gene expression in all cell types described in each analysis; users can also download the processed datasets for custom analysis. We envision that this resource will enable our community to develop and test novel hypotheses, glean more insights about cardiac development by applying innovative computational algorithms and advance our understanding of how the early events of cardiogenesis are orchestrated.

The extant cardiac development literature focused on cardiac lineage determinants and cell-type-specific marker genes was invaluable for defining the identities of the cell clusters revealed by our unbiased clustering approach. This exercise, which was critical for downstream analysis and interpretation of the data, was relatively

straightforward and is a testament to the vast and rigorous body of knowledge that has been compiled by the field. Given that we were able to identify all previously classified cardiac cell populations that we would expect to be present at these embryonic time points, we believe that our dissociation protocol did not bias us against capturing any known cardiac cell types. We however cannot eliminate the possibility that small populations were missed due to technical issues, such as dissociation bias or limited RNA capture inherent to the 10X Genomics platform, which precludes their identification during the analysis. It may also be the case that rare but distinct cardiac progenitor and differentiated populations were missed due to our computational analysis strategy. Application of rare cell type detecting analysis pipelines such as Giniclust[133] and FiRE[134] may reveal the subtle transcriptional signatures of these proportionally minor, but likely functionally important, cell subsets.

The AHF and pSHF progenitor domains comprise multipotent cell subsets that give rise to myocardial, endocardial and smooth muscle lineages of distinct regional compartments of the heart[11,46]. The additional heterogeneity revealed in our CPC reclustering analysis may represent these distinct sublineages. Moreover, these clusters may correspond to distinct medial-lateral or anterior-posterior positions within the AHF and pSHF. We attempted to gain further insights into this latter question using fluorescence *in situ* hybridization. These efforts were not fruitful due to the technical limitations of several mRNA probes not yielding signal. Sequencing-based modalities that capture both spatial and transcriptional information at high resolution[135–137] are rapidly becoming widely employed, and are more suited to address this question. We also identified heterogeneity in the endoderm overlaying the cardiac mesoderm at E7.75. It is

intriguing to consider whether the spatial positioning of endoderm subpopulations relative to the adjacent cardiac mesoderm determines specification of specific cardiac subtypes; this hypothesis can also be readily tested with spatial transcriptomics.

Our comparison of LV and RV myocardium gene signatures revealed that very few transcriptional features can distinguish between the two cell types. This is underscored by the emergence of a single ventricle cluster in our myocardium analysis and the need to recluster this population in isolation to reveal the true ventricular cell heterogeneity. This reclustering analysis yielded a distinct third population that appeared to be less differentiated early RV myocardial cells; the presence of this precursor population for the RV but not LV cells likely reflects the delayed differentiation kinetics of SHF progenitors compared to FHF progenitors[28]. We identified strong enrichment of *Cck* in the RV and IVS, but not LV cells. It is unclear what role *Cck* plays during cardiogenesis in these cell compartments as *Cck* knock-out mice are viable and fertile[138]. We did not detect any expression of the *Cck* receptor genes, *Cckar* and *Cckbr*, in the captured mesoderm populations (data not shown). A study of *Cck* processing in adult pig hearts determined that cardiac pro-Cck is post-translationally modified to a unique, triple-sulfated and N-terminally truncated product that is not present in Cck peptides of the intestine and nervous system[139]. This uniquely processed Pro-Cck cannot bind its canonical receptors; thus, it is possible that this Pro-Cck form is also expressed in cardiogenesis and its function relies on other receptors that remain to be uncovered.

The enrichment of *Cck* in the RV and IVS cells can be utilized as a tool to further probe the biology and developmental regulation of these cell types. For example, we could leverage high-resolution live imaging technologies such as light sheet microscopy

to visually map and characterize the behavior of *Cck*+ lineages labelled by constitutive or spatiotemporally controlled *Cck*-cre expression. These endeavors will shed light on the cellular dynamics of RV and IVS morphogenesis. Furthermore, in the next chapter, we demonstrate the utility of *Cck* expression as a marker that distinguishes RV from LV myocardium for dissecting the *Hand2*-null developmental phenotype.

Taken together, this work presents a key resource of cardiac single-cell transcriptomes and an opportunity for broadening and refining our understanding of the critical early stages of heart formation. This foundation will be built upon in combination with innovative emerging single-cell technologies that measure complementary molecular features to generate a comprehensive regulatory roadmap of early cardiogenesis.

Chapter 3

Analysis of *Hand2*-deficient cardiogenesis

3.1 Rationale

Hand2 is a bHLH transcription factor that is broadly expressed in multiple cardiac lineages, including the SHF-derived myocardium, endocardium and cardiac neural crest-derived mesenchyme[77]. Global loss of *Hand2* results in embryonic lethality by E10.5 due to a severely hypoplastic right ventricle and a shortened outflow tract[11]. The molecular mechanisms by which *Hand2* loss leads to dysregulation of the SHF, particularly RV and OFT progenitors, are unclear. Thus, the second goal of my thesis was to dissect how global *Hand2* deletion causes dysregulation in distinct sublineages of the SHF and leads to a heart defect. Additionally, our lineage-specifier analysis of single-cell transcriptome data from normal cardiogenesis identified *Hand2* as a determinant of the OFT myocardium fate from the AHF. To resolve the discrepancy of the predicted lineage-specifying function of *Hand2* in the OFT, but not RV, and the morphologic loss-of-function consequence, we investigated the effects of *Hand2* deletion on the fate and behavior of AHF progenitors using single-cell analyses.

3.2 Transcriptional perturbation of *Hand2*-null cardiac populations

We captured single-cell transcriptomes from *Hand2*-null embryos at E7.75, E8.25 and E9.25 and compared these data to somite-matched WT controls. Given that the *Hand2*-null mutants and WT embryos are morphologically indistinguishable at E7.75 and E8.25, while the mutant hearts are clearly malformed at E9.25, we processed the transcriptomes of WT and *Hand2*-null cardiac cells captured at the two earlier stages together, and analyzed the E9.25 transcriptomes independently (**Fig. 3.1**; **Table 3.1; Supplementary Table 3.1**).

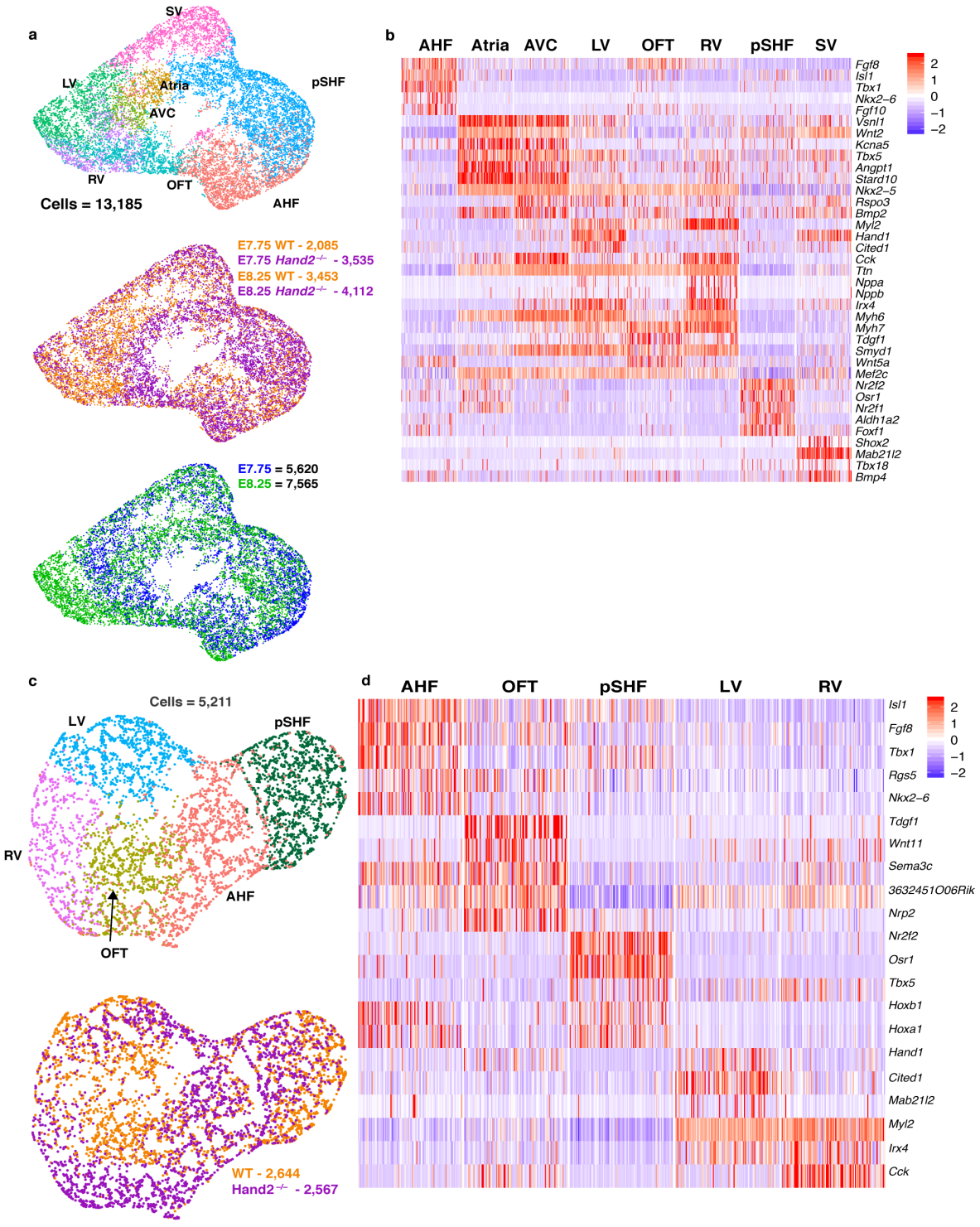

Figure 3.1: Capture of cardiac transcriptomes from Hand2-null embryos.
a, UMAP plot of WT and *Hand2*-null cells colored by cluster, genotype and embryonic stage. **b**, Heatmap of marker genes of populations from **a**. **c**, UMAP plot of cardiac populations captured at E9.25 colored by cluster and genotype. **d**, Curated list of marker genes in cardiac populations at E9.25. Scale indicates Z-scored expression values.

We used the same analysis pipeline that was employed for clustering the transcriptomes from WT embryos with one notable difference: when regressing out sources of technical variability that could affect clustering, we also included the *Hand2* gene as a potential confounding variable. This analysis method resulted in clusters that comprised both WT and *Hand2*-null cells (**Fig. 3.1**). We performed differential gene expression analysis between WT and mutant cells of each population and found that the AHF, OFT and RV precursors were transcriptionally dysregulated as early as E7.75, well before visible manifestation of the cardiac defect (**Fig. 3.2a-m; Supplementary Table 3.1**). For example, the *Rgs5* gene, which we identified as a marker of the SHF and OFT from our WT analysis, was downregulated in *Hand2*-null AHF and OFT cells at E7.75 with exacerbated dysregulation at E8.25 (**Fig. 3.2a, c, l, m**). The chromatin remodeling gene *Smyd1* was downregulated in *Hand2*-null AHF cells at E8.25, consistent with the observation that *Smyd1*-null mutants also display a hypoplastic right ventricle (**Fig. 3.2f**)[140]. *Smyd1* has been shown to regulate *Hand2* expression, and our data suggests a positive feedback regulatory mechanism between these two genes. Other cardiac genes that were dysregulated in the AHF upon *Hand2* loss included *Cfc1, Lefty2, Bves*, *Popdc2* and *Nebl*[141–144].

The retinoic acid binding protein-encoding genes, *Crabp1* and *Crabp2* emerged as highly dysregulated in the AHF, OFT and RV of *Hand2*-null mutants at E7.75 (**Fig. 3.2a-c**). These genes are known to be opposing regulators of RA signaling, which defines and patterns the posterior second heart field progenitor compartment. *Crabp1* binds and sequesters RA in the cytoplasm for degradation by RA catabolizing enzymes, while *Crabp2* binds it and promotes its nuclear import to facilitate RA-mediated transcriptional

activation[124,145]. In *Hand2*-null AHF cells at E7.75, *Crabp1* was downregulated while *Crabp2* was upregulated, leading us to hypothesize that RA signaling is ectopically activated in the *Hand2*-null AHF. To test this hypothesis, we generated mice heterozygous for the *Hand2*-null allele and a *LacZ* reporter transgene under the control of a retinoic acid response element (RARE-*LacZ*). We intercrossed these mice and examined the distribution of *LacZ* expression in the resulting embryos by *in situ* hybridization. Remarkably, the *LacZ* reporter was ectopically expressed in the AHF progenitors extended into the cardiac outflow tract region in the *Hand2*-null mutants, while being restricted to the pSHF progenitor domain in WT embryos, as expected (**Fig. 3.2j, m**). We also observed upregulation of *Hoxa1* and *Hoxb1*, which are established RA transcriptional targets and pSHF markers[55], in the AHF at E8.25 and E9.25 (**Fig. 3.2h, k, m**), as well as the pSHF gene *Osr1* at E7.75 (**Supplementary Table 3.1**) lending further support for our hypothesis of AHF posteriorization. Moreover, the gene *Upp1*, which is typically expressed in the posterior second heart field derivatives, such as the atria, was ectopically expressed in *Hand2*-null OFT and RV cells, indicating that these populations were also posteriorized (**Fig. 3.2a, b, d, e, l, m**). This suggests that while the AHF is mispatterned in *Hand2*-null embryos, transcriptional features of this abnormal identity are also conferred to the daughter cell types that differentiate from it. Overall, these observations suggest that in *Hand2*-null mutant hearts, the initial SHF lineage commitment to the AHF and pSHF occurs, but AHF and AHF-derived cells exhibit aberrant transcriptional features of the pSHF and its derivatives.

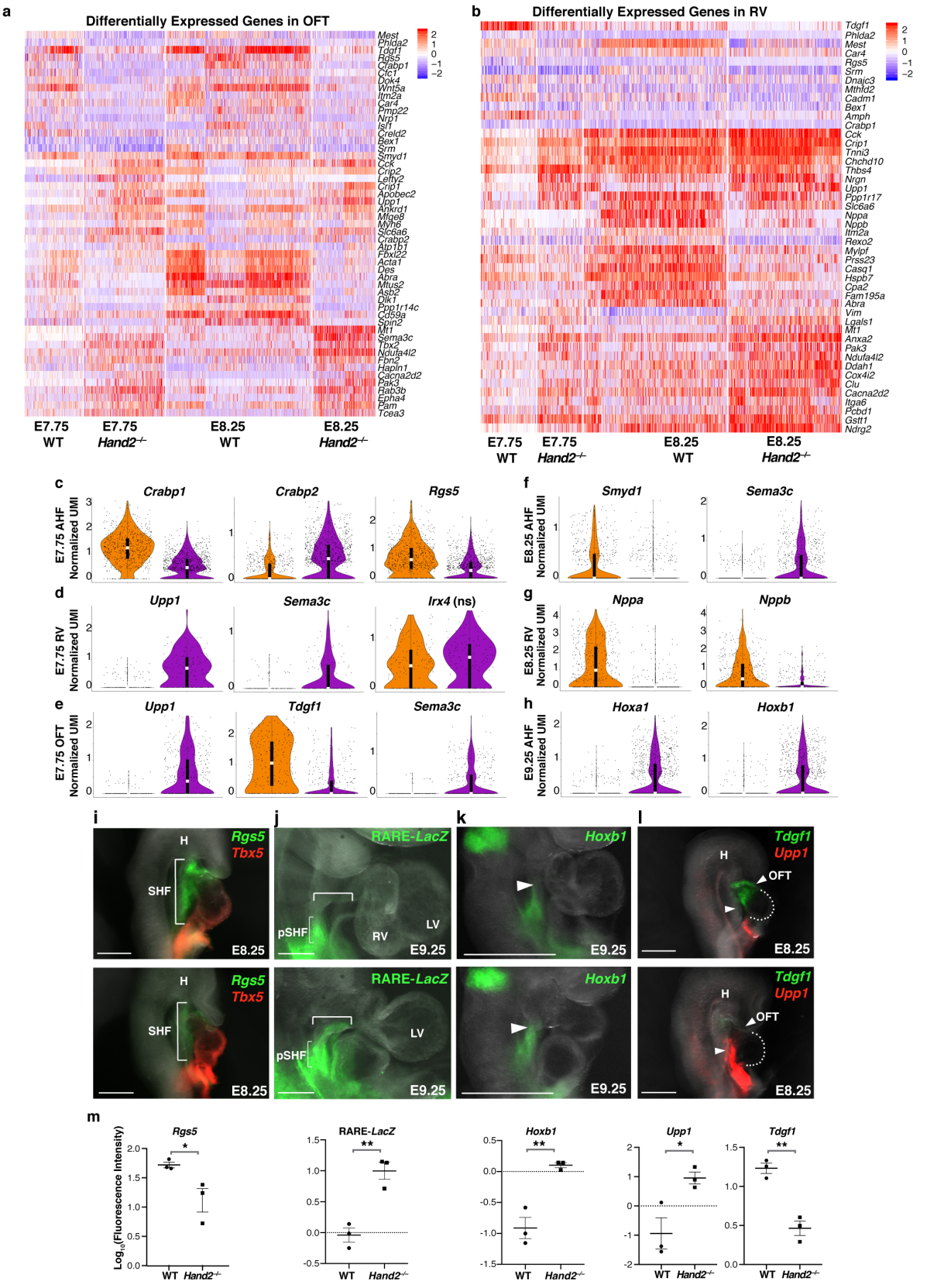

Figure 3.2 legend continued on next page

Figure 3.2: Transcriptional perturbation of *Hand2*-null cell populations.
a, Heatmap of differentially expressed genes between WT and *Hand2*-null OFT and **b**, RV cells captured at E7.75 and E8.25. Scale indicates Z-scored expression values. **c-h,** Violin plots showing dysregulated gene expression in the indicated cell compartments at indicated stages. **i-l**, mRNA *in situ* hybridization for expression of *Rgs5* and *Tbx5* at E8.25 (**i**), *RARE-lacZ* transgene (**j**) and Hoxb1 (**k**) at E9.25, and *Tdgf1* and *Upp1* at E8.25 (l). m, Quantification of fluorescence signal for indicated genes in **i-l.** n=3 independent embryos per genotype. The mean +/- s.e.m is indicated.

In addition to the posteriorized and delayed progenitor gene signature, the AHF and its derivatives displayed broader dysregulation. The gene *Sema3c* was ectopically expressed in the *Hand2*-null AHF at E8.25, and OFT and RV progenitors as early as E7.75 (**Fig. 3.2a, d-f**) *Sema3c* is an indicator of OFT cell differentiation and function, and is typically expressed later in the OFT myocardium to attract *PlexinA2*-positive neural crest cells at E8.5-E9.0 for OFT morphogenesis and septation[24]. This suggests that the regulatory mechanisms that prevent premature *Sema3c* expression in the AHF and its derivatives are abrogated upon *Hand2* loss. The gene *Tdgf1*, which is highly restricted to the OFT at E8.25, was downregulated in *Hand2*-null OFT cells at E7.75 and almost absent in the OFT at E8.25 (**Fig. 3.2a, e, l, m**). Notably, WT and *Hand2*-null RV cells had equivalent expression of *Irx4* at E7.75, suggesting that *Hand2* loss did not prevent acquisition of a ventricular identity (**Fig. 3.2d**). However, *Nppa,* a known *Hand2*-target gene[146], and *Nppb*, two markers of differentiation towards working myocardium were downregulated in *Hand2*-null RV cells at E8.25, indicating that these cells were dysregulated after specification (**Fig. 3.2g**).

In order to determine whether the posteriorization of *Hand2*-null AHF progenitors resulted in additional defects related to AHF cell proliferation or differentiation, we calculated the proportion of WT and *Hand2*-null cells in each population that were

captured at E7.75. This analysis revealed that AHF cell numbers were not decreased in *Hand2*-null embryos (*n=5*), but rather both AHF and pSHF cell numbers were higher in the mutants than in WT embryos (*n=3*) (**Fig. 3.3a**). This increase in *Hand2*-null AHF and pSHF cells did not appear to be due to differences in proliferation because for each population, relatively equivalent numbers of cells in each cell cycle stage were captured from WT and *Hand2*-null embryos (**Fig. 3.3b**). However, at E9.25, WT AHF cells were efficiently activating genes involved with cardiac muscle contraction and sarcomere organization, such as *Tnnt2* and *Actc1*, indicating that these cells were differentiating to myocardium. In contrast, *Hand2*-null AHF cells retained high expression of progenitor genes such as *Tbx1*[147] and *Fgf8*[117], which are typically downregulated as the AHF differentiates to its OFT and RV derivatives (**Fig. 3.3c**). Similarly, the *Hand2*-null pSHF compartment exhibited defects in differentiation at E8.25 (**Fig. 3.3d**). These data suggested that *Hand2*-null AHF and pSHF cells remained in a progenitor state unlike their WT counterparts which continued to differentiate.

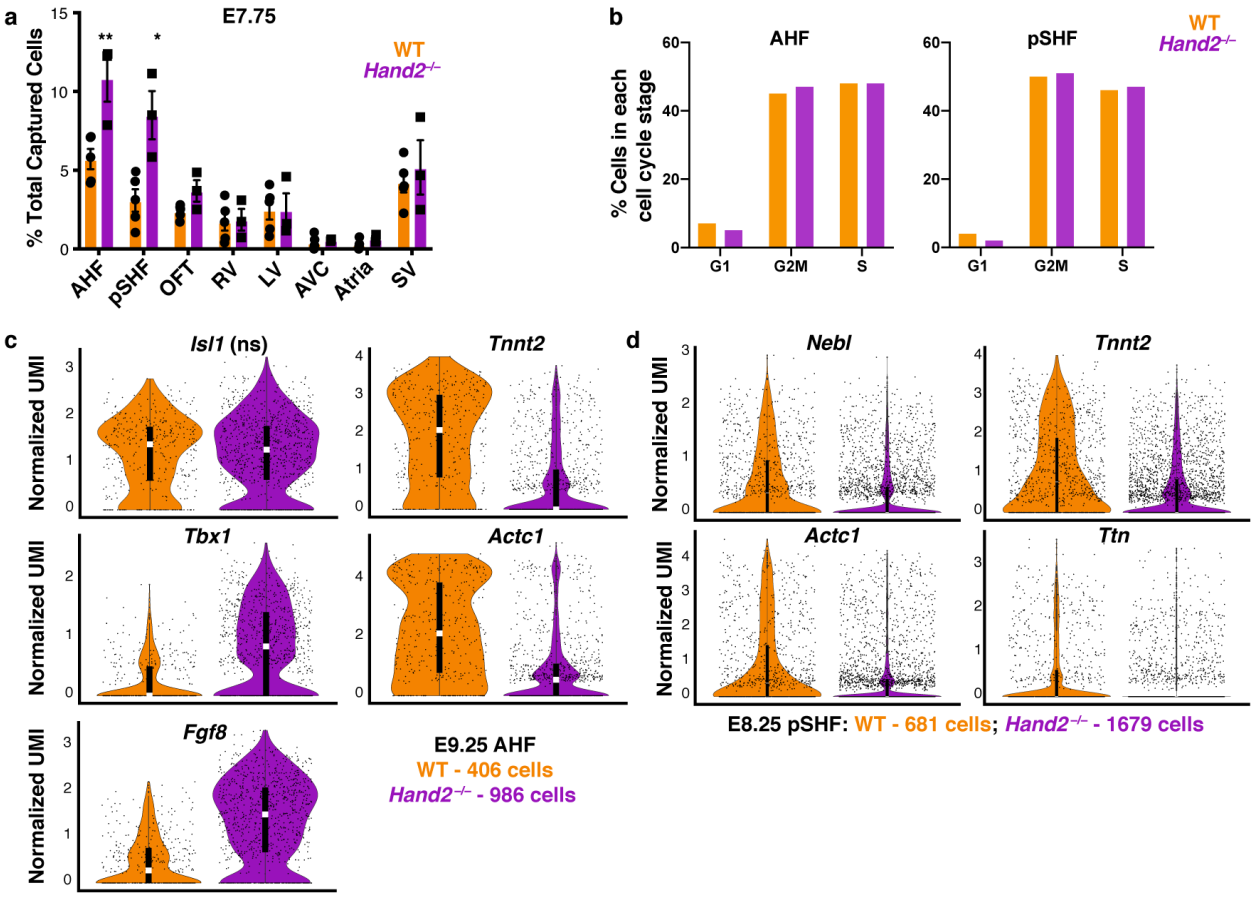

Figure 3.3: *Hand2*-null SHF cells exhibit differentiation defects.
a, Proportion of WT and *Hand2*-null cells from each population captured at E7.75. **b**, Proportion of WT and *Hand2*-null AHF and pSHF cells in each cell cycle stage at E7.75. **c, d,** Differentially expressed genes indicating a differentiation defect in the AHF (**c**) and pSHF (**d**) of *Hand2*-null embryos.

Hand1 is partially redundant with *Hand2* in several contexts including the LV[77], which is consistent with the intact LV in *Hand2*-null embryos at E9.25. We found that *Hand1* is upregulated in the *Hand2*-null LV progenitors at E7.75 (**Supplementary Table 3.1**); it is possible that this upregulation is a compensatory mechanism in response to *Hand2* loss. *Hand1* is also co-expressed in OFT progenitors at E9.25 but fails to compensate for *Hand2* loss in this cell type. Intriguingly, we found that *Hand1* was selectively downregulated in *Hand2*-null OFT, but not LV, progenitors, suggesting that the regulation of *Hand1* expression in the LV and OFT are differentially dependent upon *Hand2* function (**Fig. 3.4a, b**). Thus, *Hand2*-null OFT progenitors lack both *Hand1 and Hand2*, suggesting that *Hand1*-mediated compensation does not occur in this cell type. Additionally, our analysis of differentially expressed genes between WT and *Hand2*-null AHF, OFT and RV cells indicated that WT cells were enriched in genes that modulate normal cellular functions, while *Hand2*-null cells expressed genes controlling the response to hypoxia and apoptosis (**Fig. 3.4c**). At the time of dissection, *Hand2*-null embryos did not show overt signs of heart failure, such as cardiac edema (**Fig. 3.4d**), but these data indicate that mutant embryos were already in hypoxic stress.

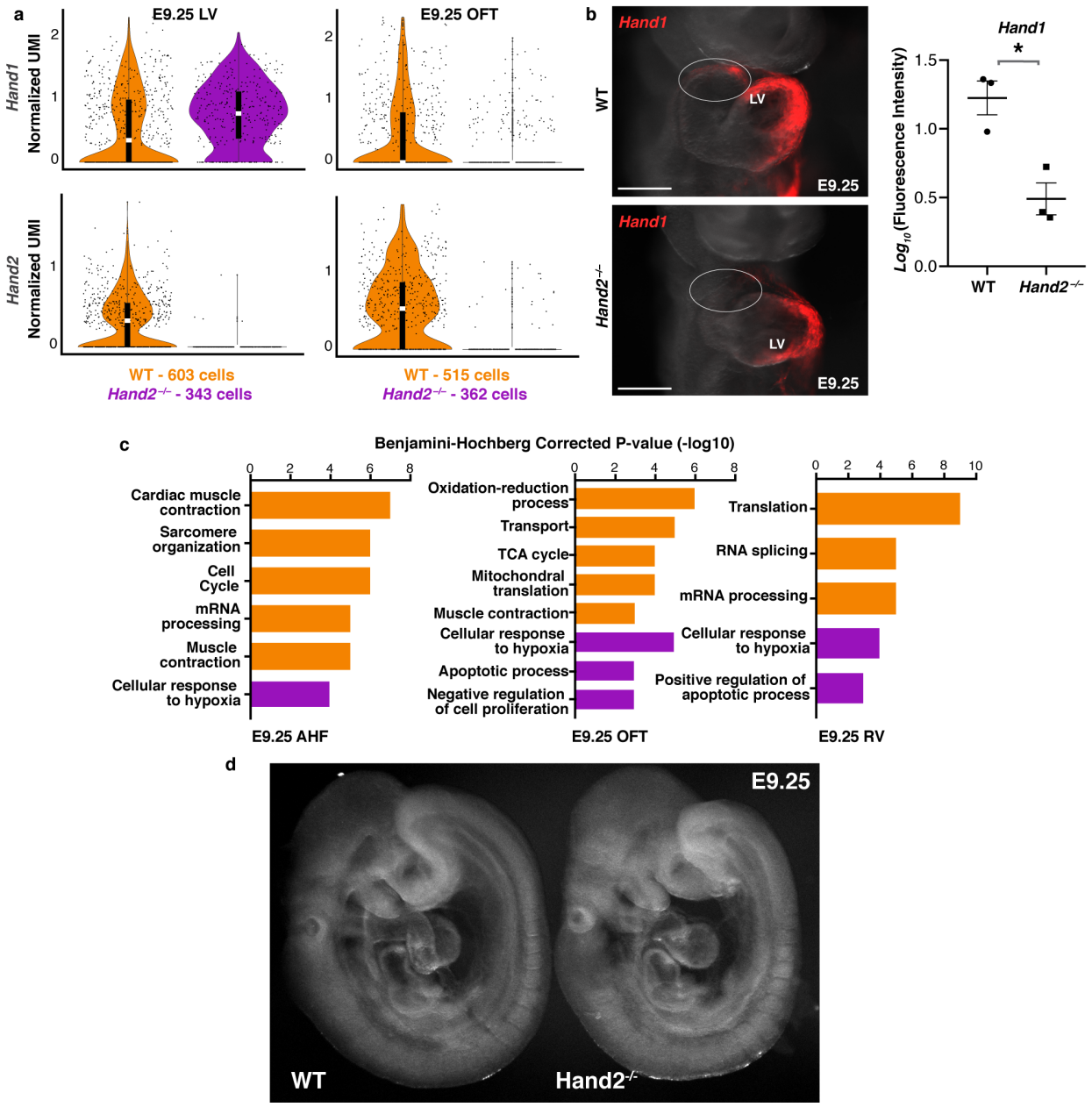

Figure 3.4: *Hand1* expression is dysregulated in OFT cells of *Hand2*-null mutants. **a,** *Hand1* and *Hand2* expression in LV and OFT cells at E9.25. **b**, *In situ* hybridization for *Hand1* in WT and *Hand2*-null embryos at E9.25 with quantification. Mean +/- s.e.m is indicated. n=3 independent experiments with similar results. **c**, GO biological process terms of differentially expressed genes in WT and *Hand2*-null AHF, OFT or RV cells at E9.25. **d**, WT and *Hand2*-null embryos used for single-cell experiments.

3.3 Pseudotime analysis of WT and *Hand2*-null AHF, OFT and RV cells at E8.25

OFT and RV progenitors are reportedly derived from different progenitor compartments of the AHF that share common ancestry with distinct head muscle lineages[100]. While lineage relationships cannot be inferred through pseudotime analysis, we reasoned that differences in cell states between WT and *Hand2*-null AHF, OFT and RV cells would indicate whether and how *Hand2* loss differentially impacts RV and OFT cell specification. Thus, to test the cell fate determinant prediction, we ordered these cells captured at E8.25 in pseudotime (**Fig. 3.5 a-d**). The resulting trajectory began with AHF cells and split into three cell states: one state primarily comprised OFT cells, while RV cells bifurcated into two separate states. *Hand2*-null cells were severely depleted in the OFT State, and while RV State 1 comprised both WT and *Hand2*-null cells in comparable numbers, WT cells predominated in RV State 2. Differential gene expression analysis between these two RV states indicated that maturation genes such as *Nppa* and *Nppb* were highly expressed in the WT-enriched State 2 (**Fig 3.5e**). These data suggest differential perturbations in OFT- and RV-fated cells upon *Hand2* loss; whereas OFT cells had disrupted specification, RV cells displayed differentiation defects but were appropriately specified, consistent with the lineage specifier analysis.

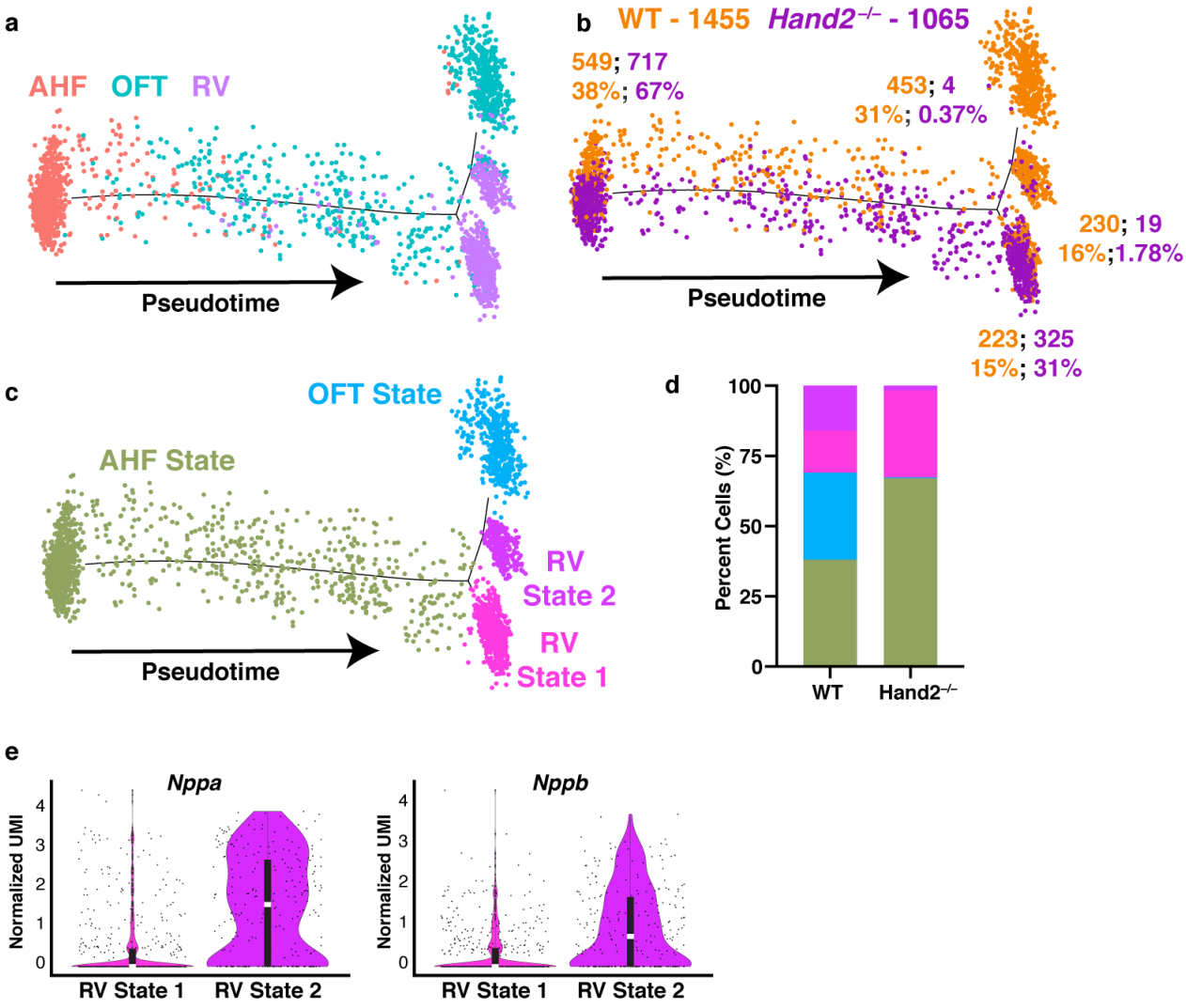

Figure 3.5: *Hand2* loss disrupts OFT myocardial cell specification and RV myocardial cell differentiation.

a, Pseudotime trajectory of AHF, OFT and RV cells at E8.25 colored by cluster identity, **b,** genotype and **c,** cell state. n=2 biologically independent embryos per genotype. Numbers in **b** indicate absolute number of cells of each genotype in each cell state. Percentages indicate proportions of cells per genotype in each state, graphically represented in (**d**). **e**, Violin plots showing expression of *Nppa* and *Nppb* each RV State. Statistics for differential gene expression tests were applied to n = 251 cells per State.

3.4 *Hand2*-null RV cells are present at E9.25 and have defects in migration

We next asked whether RV cells are retained in *Hand2*-null embryos at E9.25. Indeed, expression overlap between the ventricle gene *Irx4* and the RV-enriched *Cck* gene, as well as exclusion of LV genes *Hand1* and *Cited1*, indicated the presence of an RV population that comprised *Hand2*-null cells in comparable numbers to WT (**Fig. 3.6a, Table 3.1**). This suggested that RV cells are present in *Hand2*-null hearts at this stage, despite the absence of the RV chamber. To determine where these RV cells were located in the embryo, we performed multiplexed *in situ* hybridization for *Irx4* and *Cck* transcripts on whole-mount embryos at E8.5 and E9.25. These experiments revealed that *Irx4*$^{+}$*/Cck*$^{+}$ RV cells were present and located in the AHF area behind the LV at E8.5 and in the area of the OFT at E9.25 (**Fig. 3.6b, c**). We further validated this result using section analysis and *in situ* hybridization expression analysis for the gene *Sema3c* in whole-mount embryos at E8.5 (**Fig. 3.6d, e**). Ectopic activation of *Sema3c* in cardiac neural cells has been shown to abrogate their migration into the anterior pole of the heart. The upregulation of *Sema3c* and dysregulation of migration-related genes, such as *Ifitm1*[148], *Cxcl12*[149]*, Syne2*[150] and *Pdpn*[151], in *Hand2*-null RV cells at E7.75 and E8.25, suggested a migratory defect as the underlying mechanism for the mislocalization of these cells at E9.25 (**Supplementary Table 3.1**). Moreover, the genes *Wnt5a* and *Tbx2* were downregulated and ectopically activated, respectively, in *Hand2*-null RV progenitor cells at E7.75 and E8.25 (**Fig. 3.6f**). Deletion of *Wnt5a* and misexpression of *Tbx2* in the AHF have previously been shown to inhibit deployment of AHF cells into the heart tube[152,153], suggesting their dysregulated expression in *Hand2*-null RV cells may also contribute to a migratory defect.

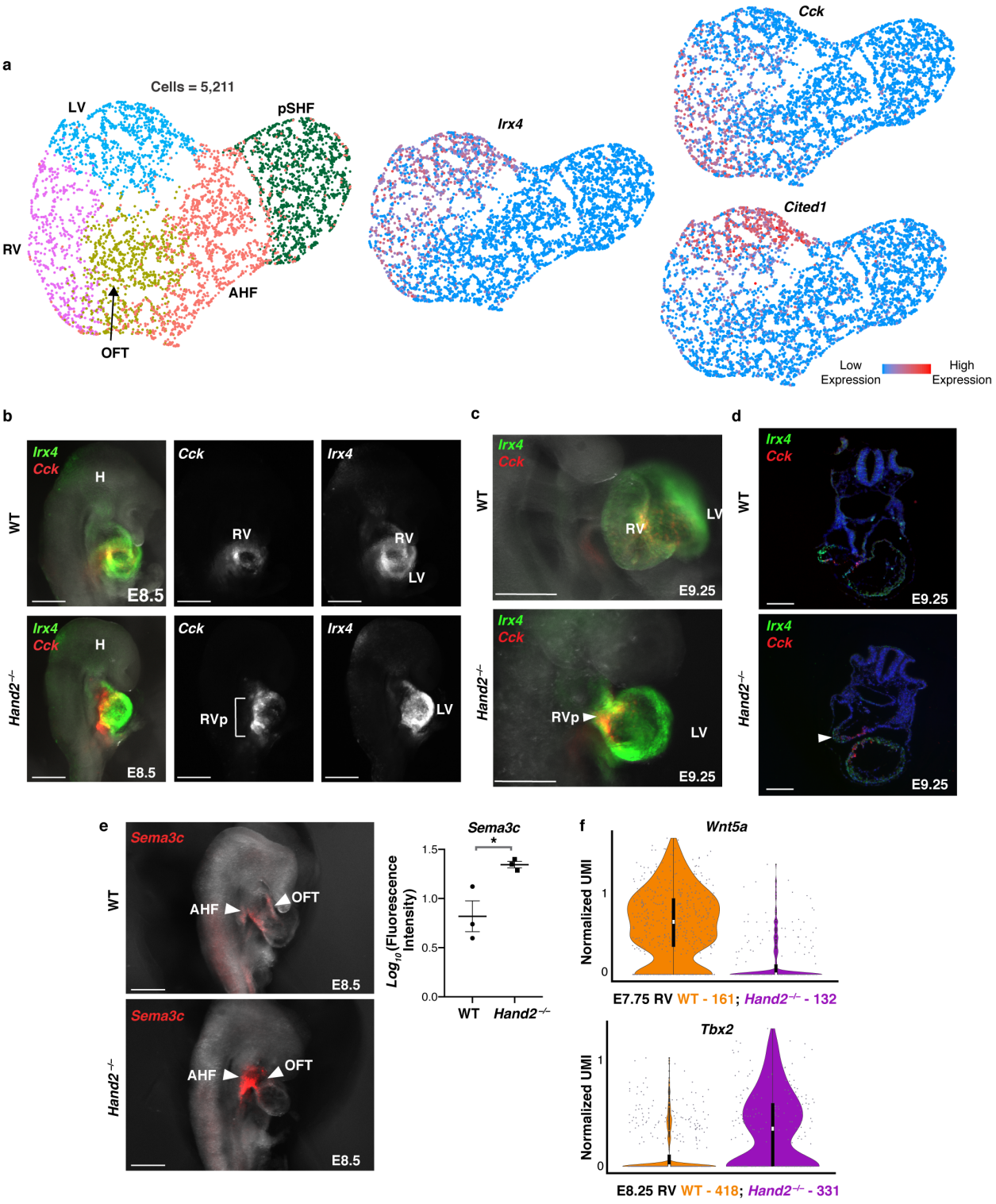

Figure 3.6: Right ventricle cell migration is impaired in *Hand2*-null embryos.
a, UMAP plot of subset of cardiac populations captured at E9.25 colored by cluster and showing expression domains of *Irx4, Cck* and *Cited1* indicating the presence of LV and RV cells. **b,** Expression of *Cck* and *Irx4* by *in situ* hybridization in right lateral view to

localize RV cells at E8.5 and corresponding fluorescence image for *Cck* or *Irx4* expression alone. **c**, Whole-mount *in situ* hybridization for *Irx4* and *Cck* in right lateral view and **d,** transverse sections at E9.25 indicating presence of RV cells in *Hand2* mutants (arrowheads). n=2 independent experiments with similar results for **b-d**. **e**, Expression of *Sema3c* to localize AHF cells and derivatives at E8.5 in right lateral view with quantification of signal. n = 3 independent experiments with similar results. Mean +/- SEM is indicated. Scale bars, 200 μm. **f**, Violin plots of *Wnt5a* and *Tbx2* expression in WT and *Hand2*-null RV cells at E7.75 and E8.25, respectively.

Our *in situ* hybridization strategy to identify *Hand2*-null RV cells with *Irx4* and *Cck* probes revealed that these cells were located behind the LV in the area of the OFT region at E9.25 (**Fig. 3.6c**). Given that *Cck* strongly marks the interventricular septum (IVS), the striking lack of *Cck* expression along the boundary of LV chamber, where the normal IVS should morphologically be situated, also indicated that IVS formation was disrupted. This is consistent with studies showing that *Hand2* function is critical for IVS development[154].

3.5 Conclusions and Discussion

In this work, we interrogated how loss of *Hand2* disrupts the fate and behavior of the anterior heart field and its derivatives using single-cell RNA sequencing. We found that transcriptional dysregulation of *Hand2*-null cardiac progenitors occurs as early as E7.75, concomitant with the onset of *Hand2* expression in the cardiac crescent and prior to emergence of a visible morphological defect. To our knowledge, this is the first report of the *Hand2*-null phenotype that investigates the dysregulation of cell types at this early stage. The *Hand2*-null AHF and pSHF populations both displayed delayed differentiation kinetics, with the AHF displaying aberrant transcriptional features of the pSHF that appears to be mediated by anterior expansion of retinoic acid signaling and subsequent posteriorization. Strikingly, consistent with the predictions of our lineage-specifier analysis

(Chapter 2), we uncovered that loss of *Hand2* perturbs RV and OFT myocardium cells in distinct ways; OFT cells fail to become specified while RV cells are appropriately specified, but fail to further differentiate and migrate into the heart tube to eventually form the RV chamber.

Since its discovery over twenty years ago, *Hand2* has been considered a crucial RV gene, due to the dramatic loss of the RV chamber upon its deletion. While our findings indicate that *Hand2* is indeed critical for RV myocardium differentiation and migration, the discovery that *Hand2* is important for OFT myocardium specification was previously unappreciated. While the important contribution of *Hand2* to the formation of the OFT structure has been established[80,81], the current literature on this has largely focused on *Hand2* function in cardiac neural crest cells that migrate into the heart to regulate morphogenesis and septation of the OFT. Our work further refines our understanding of *Hand2* function in these distinct progenitor compartments and sheds light on the molecular mechanisms by which *Hand2* loss leads to a heart defect.

A previous study reported ChIP-seq data collected from whole hearts of Hand2^{3xFLAG} mouse embryos, and defined the genome-wide interaction profile and putative transcriptional targets of *Hand2*[85]. Many of the genes that neighbored the regions bound by *Hand2* in this study were transcriptionally dysregulated in the AHF, RV and OFT cells in our dataset, including *Lefty2, Pitx2, Wnt5a, Tgfb2, Bmp4, Smyd1, Mtus2* and *Rgs5.* This provides further support for these genes being bona fide *Hand2* targets. Other genes that were not identified in this study but were strongly dysregulated in our dataset include *Tdgf1, Upp1, Sema3c, Cfc1* and *Nppa. Hand2* has previously been shown to regulate the *Nppa* gene, but this regulation occurs independently of its DNA-binding

activity[146]. It is possible that strongly dysregulated genes such as *Tdgf1*, *Sema3c* and *Upp1* are similarly regulated by *Hand2*.

Several genes encoding regulators of left-right asymmetry were dysregulated in *Hand2*-null AHF, pSHF, OFT and LV, including *Nodal,* its co-receptor *Cfc1, Lefty2* and *Pitx2* (**Supplementary Table 3.1**). Intriguingly, the direction of transcriptional dysregulation of *Pitx2* in distinct compartments was variable; *Pitx2* was downregulated in the mutant pSHF at E7.75, consistent with the downregulation of *Nodal* in this domain, but was upregulated in the mutant OFT and LV at the same stage. Moreover, *Lefty2* was upregulated in the mutant AHF and OFT at E7.75, while *Cfc1* was downregulated in all AHF, pSHF, OFT and LV compartments. Early work on the relationship between *Hand2* and establishment of L/R asymmetry suggested that *Hand2* function in the ventricle chambers was decoupled from L/R patterning[78]; our single-cell data indicates that *Hand2* loss does impact expression of mediators of L/R asymmetry signals in the SHF and OFT. Whether this transcriptional perturbation reflects purely dysregulated expression of these genes as opposed to disruption of the cell populations that express these genes is unclear. In an attempt to distinguish between these two possibilities, we subclustered the AHF and pSHF populations of WT and *Hand2*-null embryos to study the subcompartments within these domains; however, mapping the *Hand2*-null populations to their WT counterparts was not trivial or fruitful (data not shown). We are currently in the process of addressing this with computational methods such as transfer learning[155,156]. Additionally, defining how the spatial distribution of these populations is affected in *Hand2*-null mutants will also yield insights into this question.

We observed gene dysregulation in both the AHF and pSHF populations at E7.75, however the onset of overt transcriptional perturbation in their respective daughter lineages was not synchronous. The AHF-derived OFT and RV myocardium cell transcriptomes were dysregulated at E7.75 in *Hand2*-null mutants, while pSHF-derived atrial cells became dysregulated at the heart tube stage at E8.25 (**Supplementary Table 3.1**). One possibility for this observation is that the transcriptional consequences of *Hand2* loss in the pSHF are delayed relative to AHF, and might reflect subtle differences in timing of *Hand2* expression onset in the cardiac crescent. In support of this interpretation, our wild-type analysis of CPC populations indicates that at E7.75, the expression levels of and proportion of cells expressing *Hand2* are higher in the AHF than the pSHF (**Fig. 3.7a)**.

While we did not observe overt signs of heart failure during dissection of *Hand2*-null mutants at E9.25, these embryos were in hypoxic stress based on enrichment of genes related to the response to hypoxia and positive regulation of apoptosis. However, at earlier stages, we did not find evidence of increased apoptosis, neither transcriptionally nor via an increase in contaminating mitochondrial reads[157]. Although we cannot rule out the fact that *Hand2* regulates the survival of SHF progenitors, as observed previously at E9.25, especially given that cell death pathways are primarily regulated at the post-translational level[158], our transcriptome data suggests that at early stages of cardiogenesis, *Hand2* loss does not dramatically impact cell survival.

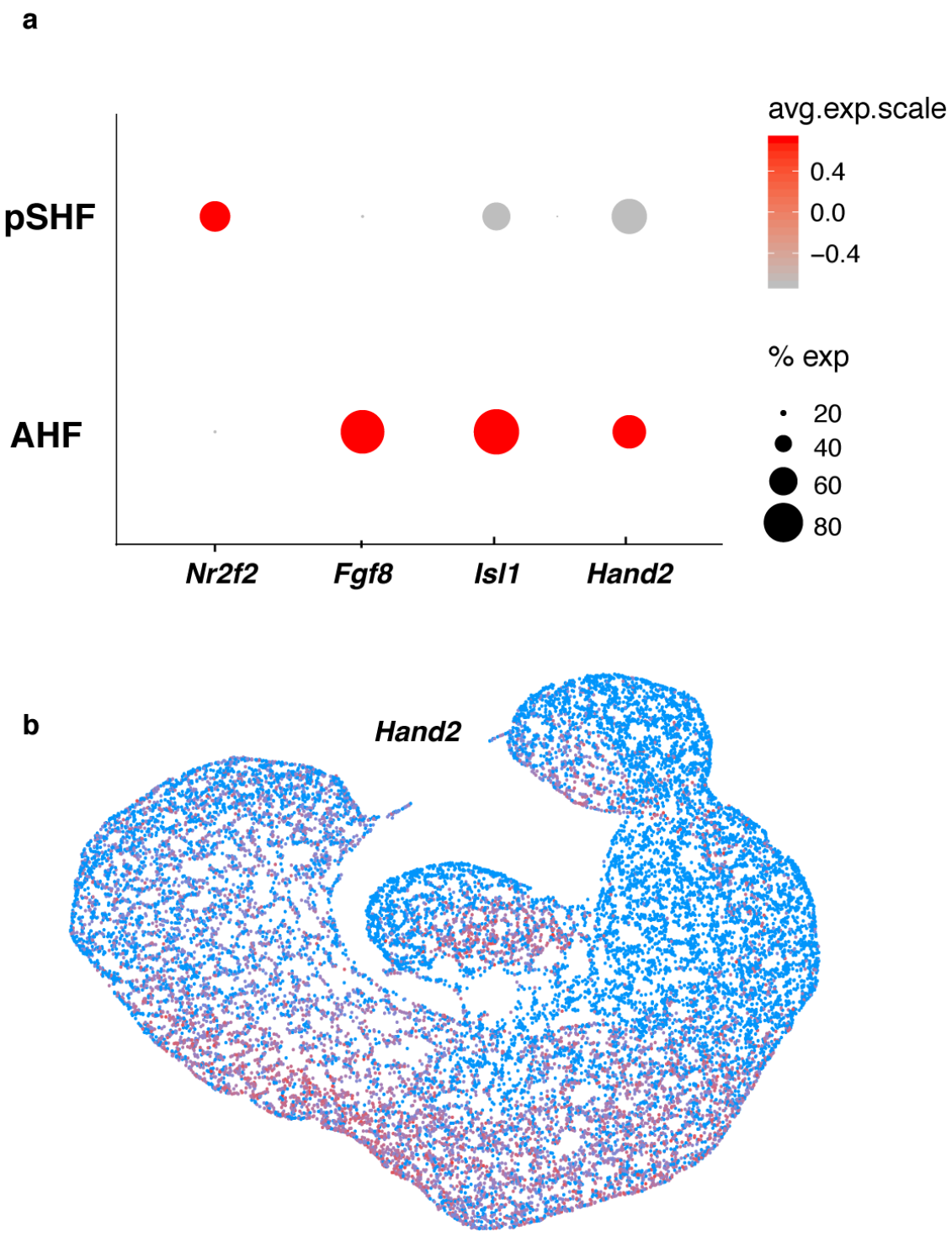

Figure 3.7: Enriched expression of Hand2 in AHF and broad expression in the cardiogenic mesoderm.
a, Dotplot indicating expression of Hand2 in the AHF and pSHF populations at E7.75. **b**, UMAP plot indicated broad expression of Hand2 in cardiogenic mesoderm.

RA abundance and signaling gradients in the developing embryo are governed by cell-type-specific enzymatic networks that promote its metabolism and subsequent transcriptional signaling or limit its bioavailability by increasing its catabolism[159,160]. Our data indicates that *Hand2* loss results in transcriptional dysregulation of genes encoding two RA binding proteins with opposing functions and impacts on RA signaling. The pattern of dysregulation of these genes suggested that RA signaling is expanded into the *Hand2*-null AHF and OFT. Indeed, we validated the anterior expansion of RA with a reporter of its transcriptional activity and showed anterior expression of one of its target genes, *Hoxb1*, in the SHF, indicating that the *Hand2*-null AHF domain is posteriorized.

It is possible that expanded RA signaling in *Hand2*-null AHF compartments that give rise to the OFT and RV myocardium cells might contribute to the defects we observed. This is particularly relevant given that *Crabp1* and *Crabp2* expression is also dysregulated in the OFT and RV myocardium cells at E7.75 (**Supplementary Table 3.1**) and treatment of mouse embryos with exogenous RA at E7.5 abrogates proper formation of the OFT and ventricles[51]. Intriguingly, more recent work demonstrated that exogenous RA administration does not impact ventricular specification at the cardiac crescent stage, but alters their differentiation[161], consistent with our findings in the *Hand2*-null mutant.

In order to determine whether blocking the ectopic RA activity in *Hand2*-null mutant embryos would reverse or mitigate the *Hand2*-null heart defects, we treated embryos resulting from a *Hand2*$^{+/-}$ intercross with the pan-retinoid receptor antagonist, AGN193109[162] by oral gavage administered to the pregnant female when embryos were approximately at the E7.5 stage. While we found that some *Hand2*-null embryos treated with the antagonist exhibited an enlarged OFT region and some dampening of expanded

Upp1 expression (data not shown), these phenotypes were highly variable, likely owing to the difficulty in consistent timing of antagonist administration between different litters. Stringent dissection of whether and how ectopic RA drives the defects will require complex genetic crosses. We also cannot rule out that there may be RA-independent, *Hand2*-dependent dysregulation in the AHF, since the expanded expression domain of the RA-transcription reporter did not encompass the entire AHF region at E9.25. One possibility is that *Hand2* might be directly regulating expression of *Hoxb1* and *Hoxa1*, since it has been shown to bind genomic regions in the vicinity of these two genes[85].

Several RA-dependent, Hox gene expressing progenitor subdomains of the pSHF contribute to the inferior wall of the OFT that will give rise to the myocardium at the base of the pulmonary trunk. These subdomains are defined by distinct and overlapping expression of *Hoxa1, Hoxb1* and *Hoxa3*. While *Hoxa1* and *Hoxb1* appear to be ectopically activated in the *Hand2*-null AHF, we also noted increased *Hoxa1* only in the mutant pSHF at E9.25. Whether this increase reflects upregulation in endogenous *Hoxa1*-expressing pSHF cells or that this subdomain has expanded merits further scrutiny. Furthermore, expression of *Tbx1* and *Tbx5* were upregulated and downregulated, respectively in the pSHF. A potential explanation for this observation is that decreased *Tbx5* in the mutant pSHF results in failure of *Tbx1* downregulation in cells at the AHF-pSHF boundary[57]. Further interrogation of the molecular and spatial heterogeneity of the *Hand2*-null pSHF will clarify whether this dysregulation occurs in the same cell subpopulation. Taken together, these data suggest that pSHF patterning may also be disrupted in *Hand2*-null mutant embryos.

We have shown that loss of *Hand2*, which is broadly expressed in cardiac mesoderm derivatives (**Fig. 3.7b**), causes dysregulation of distinct molecular pathways and therefore drives unique functional deficits in cell subpopulations of the developing heart. Understanding how *Hand2* molecularly regulates these pathways in OFT and RV myocardium requires a thorough characterization of its target gene network in each of these cell types. This is now technically feasible with the rapidly expanding single-cell toolkit, which includes single-cell ATAC-seq, ChIP-seq and protein-DNA interaction mapping technologies[163–165]. Combining the output of these technologies with single-cell RNA sequencing data, including the dataset presented here, will enable us to map the chromatin landscape and DNA binding activity of *Hand2* in specific cell subtypes and will shed light on the longstanding question of how this core cardiac transcription factor orchestrates cardiogenesis.

Chapter 4

Summary and future studies

4.1 Summary

In this work, we interrogated the transcriptional dynamics of early normal and *Hand2*-deficient cardiogenesis with scRNA-seq. We uncovered the cellular and transcriptional heterogeneity that emerges during normal cardiogenesis and demonstrate that these data can be leveraged to uncover novel cell-type-specific fate determinants. These data will serve as a valuable resource for the cardiac development community, particular when combined with other single-cell modalities, such as scATAC-seq, scGESTALT, ChIP-seq and Slide-Seq[166,167]. Additionally, by applying new computational methods for single-cell assays, our understanding of the molecular regulatory logic underlying how diverse cardiac cell types are integrated to assemble the heart will become clearer and more nuanced. Our analysis of *Hand2*-deficient cardiogenesis demonstrates that scRNA-seq is a potent strategy for dissecting how cell subsets acquire distinct functional and specification defects during disrupted organogenesis of a complex tissue. Thus single-cell assays will empower us to more effectively determine the cellular and molecular pathways underlying phenotypic presentation of congenital defects associated with genetic variation.

4.2 Leveraging multi-model single-cell analysis to study cardiac development

While scRNA-seq is a powerful methodology, focusing only on gene expression dynamics yields an incomplete view of the complexities underlying cardiac development. Given that scRNA-seq does not capture epigenetic, lineage and spatial information, efforts to combine it with other modalities that focus on these aspects will give us greater clarity on the molecular mechanisms that govern heart formation. For example, a study

that paired Fluidigm-based scRNA-seq with single-cell ATAC-seq (scATAC-seq) to analyze *Isl1*+ CPCs identified differential transcription factor binding dynamics as these progenitors bifurcated to the cardiomyocyte or endothelial cell lineage[70]. This study also demonstrated that scATAC-seq can uncover greater cellular heterogeneity than scRNA-seq alone, likely owing to the fact that ATAC-seq can identify chromatin remodeling changes in transition cell states that do not coincide temporally with transcriptional activation or repression.

Several questions remain that can be addressed by integrating scRNA-seq and ATAC-seq analysis. Differential enhancer usage confers a cell-type-specific regulatory code that is not evident at the transcriptional level[168,169]. Our analysis of LV and RV myocardium cells highlighted their remarkably similar transcriptional profiles and suggests that these two cell types are more distinct at the level of enhancer usage[170,171]. Similarly, very few genes emerged as specific to the CPC subpopulations that we uncovered. Identification of CPC-subtype-specific enhancers will enable a more thorough characterization of these populations, and will generate more refined lineage-specific tools to define and study the daughter cell-types that are specified from each domain[172]. Thus, combined scRNA-seq and scATAC-seq analysis has immense potential to expand the known repertoire of cardiac cell-type-specific features.

We and others have demonstrated the utility of constructing cellular differentiation trajectories with scRNA-seq data[173]. While these have revealed key insights about the transcriptional drivers of cell states that arise in a differentiating system, the predicted trajectories do not reflect true genetic lineage relationships. In the last several years, techniques that unite scRNA-seq and genetic lineage mapping have been developed,

which have revealed the rich and complex clonal dynamics of multi-lineage organs[174–177]. Application of these approaches to cardiogenesis will yield insights into longstanding questions about the timing of lineage segregation and origins of CPCs. For example, there are conflicting reports as to whether the heart and head muscle lineages share a common precursor[13,14,100]. Additionally, the existence of a common precursor for all FHF and SHF cells has been hypothesized, but the identity of such as progenitor has remained elusive[13,178]. Moreover, an early *Tbx5+/Mef2cAHF+* progenitor pool that contributes to the forming IVS has been identified[13]. The origin of this population, which expresses markers of FHF and SHF lineages that are mutually exclusive in other regions of the heart, is unclear. Combined single-cell lineage mapping and transcriptome analysis will allow more thorough investigation of these questions.

Another limitation of scRNA-seq is that spatial information is lost during sample preparation, given that cells must be dissociated from their native tissue context for capture of individual transcriptomes. Technologies such as spatial transcriptomics, Slide-seq and Seq-Fish[135–137] overcome these technical limitations and represent an exciting conduit of biological discovery. This advance is particularly relevant to embryogenesis, where the sample size is exceedingly small and capturing spatial information in a high-throughput manner is challenging. As discussed in Chapter 2, these methods can be applied to determine whether the spatial positioning of endodermal populations determines specification of distinct cardiac subtypes from the underlying cardiac mesoderm, and to precisely map the embryonic locations of CPC subdomains. Moreover, while its clear that myocardium subtypes from different chambers have distinct transcriptional signatures, whether there are chamber-specific endocardial, epicardial or

fibroblast subpopulations is unclear. Additionally, diverse subpopulations of non-myocyte cells that regulate critical functions and homeostasis of cardiomyocytes have been identified[179]. Spatial transcriptomics methods will be able to better define which specific non-myocyte subsets interact with distinct cardiomyocyte subpopulations. These endeavors will provide novel insights into how the many cell types of the heart converge and interact at 3D resolution.

4.3 Novel computational methods to interrogate cardiac development and fate

Innovative computational methods for single-cell analyses are rapidly being developed and have the potential to expand our understanding of the mechanisms underlying cardiogenesis. A recent study utilized scRNA-seq and cell-cell pairing analysis to define the signaling interactions between CPCs, and discovered that SHF progenitors are guided to the heart tube via *Cxcr2/Cxcr4* signaling cues emanating from FHF cells[180]. Application of similar computational strategies[181–183] will enable us to comprehensively define the range of intricate cell-cell signaling events in cardiac differentiation and morphogenesis, including interactions between the endoderm and cardiac cell subsets, neural crest cells and the SHF, epicardium and myocardium, and endocardium and myocardium[6,22,184–186]. These efforts will inform strategies to more faithfully recapitulate development in a dish and generate stem cell-derived cardiac cell types *in vitro* that are comparable to their *in vivo* counterparts.

Understanding how gene networks are established to specify and maintain the fate of cardiac cell subsets is now possible with computational tools such as TransSyn[187]. This method identifies synergistic transcriptional cores that define the identity of a given cell

subpopulation. These 'identity transcription factors (TFs)' can then be utilized to drive transdifferentiation of desired cardiac cell fates from a known starting cell type. The value of this approach is that it identifies distinct core TF gene regulatory networks that are active in cell subtypes with highly similar transcriptional signatures. In the long term, applying this transdifferentiation strategy may be a viable clinical intervention for CHD. For example, the LV and RV chambers perform physiologically distinct roles and are differentially impacted in CHD, despite having highly similar transcriptional signatures *in vivo*. The CHD called Hypoplastic Left Heart Syndrome (HLHS), in which the left ventricle, mitral valve and aortic valve are underdeveloped, comprises 1.4-3.8% of CHDs; yet despite this relatively low incidence, HLHS is responsible for 23% of cardiac deaths occurring in the first week of life[188]. Thus, there is a critical need for strategies to combat this devastating condition, and one such approach is to convert cells of the RV to take on a more LV-like identity and physiology that can support systemic circulation.

To this end, our lab is currently testing the TransSyn computational approach in collaboration with Dr. Satoshi Okawa and Dr. Antonio del Sol at the University of Luxembourg. Our preliminary goal is to convert *in vitro* human iPS-derived SHF cells and RV cardiomyocytes to FHF cells and LV cardiomyocytes, respectively. In our hands, human iPS-derived CPCs and cardiomyocytes resemble the SHF and SHF-derived RV cells, respectively. We applied the TransSyn method to our mouse embryonic LV, RV, FHF and AHF cells (chapter 2) and identified *Hand1, Cited1, Tbx20, Hmgb2, Mef2c, Ybx1* and *Btf3* as core identity TFs of LV cells, while *Hand1, Smarcd3, Gata5, Hoxb2, Morf4l2* and *Ddx5* were determined as identity TFs of FHF cells. We are currently testing a combinatorial strategy in which we infect iPS-CPCs, which have an SHF signature, with

combinations of FHF identity TFs in order to obtain functional FHF cells that can differentiate into LV cells. We are also implementing this strategy with iPS-CM, which have an RV identity, to more directly obtain iPS-CMs with an LV identity. If these efforts prove to be successful, our next step is to test the possibility of achieving this conversion *in vivo*. This work will lay the foundation for developing cardiac cell conversion strategies, with profound implications for our ability to treat CHD.

4.4 Harnessing scRNA-seq to understand and overcome CHDs

Single-cell transcriptomics enabled us to uncover how loss of *Hand2* leads to cell-type-specific perturbations resulting in a complex developmental defect. These features remained undiscovered due to relatively few cells being affected, particularly at early developmental stages, and likely were revealed by our ability to interrogate tens of thousands of cells in the heart. Previous studies of human patient cohorts have identified damaging loss-of-function mutations in *Hand2* that are associated with CHDs such as ventricular septal defects and Tetralogy of Fallot[189–192]. Our work presents a foundation for interrogating how such mutations could result in these heart malformations.

Future studies should also investigate how increased *Hand2* dosage causes congenital defects. One study discovered that duplication of the *Hand2* gene is a major cause for the limb and heart phenotypes of the chromosomal disorder partial trisomy distal 4q[193]. Additionally, our scRNA-seq analyses of the transcriptional changes induced by trisomy 21, the genetic condition also known as Down Syndrome (DS), reveal upregulated *Hand2* expression and concomitant dysregulation of genes such as *Upp1, Crabp1* and *Crabp2* (Sanjeev Ranade, Unpublished observation). This dysregulated

gene signature manifests in *Rspo3*+ AVC and OFT myocardium cells in a mouse model of DS as well in iPS-derived cardiomyocytes that originated from a human DS patient. Thus, uncovering the mechanisms by which ectopic *Hand2* expression disrupts the fate and behavior of these specific cell types will advance our understanding of how these genetic lesions cause developmental anomalies.

The scRNA-seq approach enabled us to glean novel, precise insights about the *Hand2*-null phenotype in a mouse model of cardiac development. Many core cardiac transcription factors, such as NKX2-5, GATA4, MEF2C, TBX5, HAND1, TBX1 and ISL1, are broadly expressed in overlapping and complementary cell populations of the heart, and are causative agents of numerous CHD phenotypes when mutated[6]. Dissecting the loss of function phenotypes of these critical genes with the single-cell modalities described is likely to reveal novel mechanisms underlying their biology. Moreover, constructing an integrated view of how these transcriptional regulators coordinately orchestrate cardiac morphogenesis is now possible. For example, independent loss of *Hand2, Isl1* and *Smyd1* gene function leads to abrogated formation of the RV chamber[18,140]. Investigating the convergence and divergence of the molecular mechanisms by which loss of these genes leads to this shared defect will yield a broader, more detailed understanding of how they synergistically and uniquely regulate RV chamber formation.

Ongoing whole-exome sequencing studies of CHD proband-parent trios are identifying damaging inherited variants in novel genes with unknown cardiac functions[5,194–196]. Our dataset serves as a critical resource of cell transcriptomes that can be interrogated to define the cell subtypes in which these novel genes function. Several

high-resolution catalogs of mouse and human cardiac cell types that emerge and function at later stages of heart formation are being reported[71,197–199]. Combining our dataset with these resources will yield a powerful tool that the cardiac development field is poised to wield in the quest to uncover the genetic and mechanistic underpinnings of CHD. This is a prerequisite for defining preventative approaches and postnatal intervention strategies for ongoing sequelae of these devastating defects.

Chapter 5

Materials and Methods

Animal Models

Animal studies were conducted in strict compliance with all relevant ethical regulations in the animal use protocols, UCSF animal use guidelines and the NIH *Guide for the care and Use of Laboratory Animals*. All protocols concerning animal use were approved by the Institutional Animal Care and Use Committee (IACUC) at UCSF and were accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). Transcriptomes were captured from wildtype (WT) and *Hand2*-null embryos from intercrossed C57BL/6 mice heterozygous for the *Hand2*-null allele. The sexes of all embryos used for capture of single-cell transcriptomes are listed in **Table 5.1.** Lineage tracing of Cck expressing cells was performed using Cck-Ires-Cre (JAX stock #012706)[200] and Ai14 (JAX stock #007914)[201] mice. Validation of ectopic RA signaling in the *Hand2* mutant was done by crossing the mutant line to RARE-hsp68*LacZ* mice (JAX stock #008477)[202].

Timed matings between male (8-10 weeks of age) and female (6-8 weeks of age) mice were set up where noon on the day of plug detection was considered E0.5. Pregnant females were identified by echocardiography performed at E6.5 and sacrificed to harvest embryos at E7.75, E8.25 and E9.25 for scRNA-seq and at E7.75, E8.25, E8.5 and E9.25 for whole mount and section-based in-situ hybridization experiments. Transcriptomes from at least 2 embryos were collected per embryonic stage, per genotype (**Table 5.1**). The sample sizes of embryos used for single-cell transcriptome analysis at each time point was chosen to obtain cell numbers comparable to estimated cell numbers in the cardiogenic region at each embryonic stage. Embryos were developmentally matched at each time point by somite count.

Randomization was not implemented; experimental groups were determined by genotype i.e., *Hand2* wild type embryos were compared to *Hand2*-null embryos. Covariates were not relevant to the analysis of the *Hand2*-null phenotype as the developmental defect is highly penetrant regardless of embryo sex, and the developmental stages analyzed were prior to the onset of overt heart failure. Investigators were not blinded to allocation of embryos during experiments. Blinding was not possible for the WT and *Hand2*-null embryo comparisons due to the need to match somite counts to control for developmental timing.

Embryo dissection and single-cell library generation

The entire cardiogenic region was dissected at each time point, including the SHF region that lies behind the cardiac crescent and heart tube, as well as the first and second pharyngeal arches at E9.25. Due to the small size of embryos at these stages, some surrounding tissue (indicated in **Fig. 2.1, Chapter 2**) encompassing the posterior lateral plate mesoderm (E7.75), head folds (E7.75 and E8.25), and endoderm, was microdissected to ensure complete retrieval of cardiac populations. Embryos were dissected in cold PBS (Life Technologies, CAT# 14190250), de-yolked and placed in PBS/1% FBS (ThermoFisher Scientific, CAT# 10439016) solution on ice until dissociation (approximately 3 hours). Yolk sac DNA was extracted (QuickExtract DNA Extraction Solution, Epicentre, CAT# QE09050) and used for genotyping to distinguish *Hand2* WT and *Hand2*-null embryos before further microdissection of cardiac regions at each stage. Dissected cardiac tissue was incubated in 200 µl TrypLE (ThermoFisher Scientific, CAT# 12563029) for 5 min, triturated with a 200 µl pipette tip, and incubated for an additional

5 min. The TrypLE solution was quenched with 600 µl PBS/1% FBS. Cells were filtered through a 70 µm cell strainer (BD Falcon, CAT# 08-771-2), centrifuged at 150 rcf for 3 min, and resuspended in 35 µl PBS/1% FBS. Single-cell droplet libraries from this suspension were generated in the 10X Genomics Chromium controller according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagent Kit v2 User Guide. The cell capture efficiency of the Chromium controller is ~57%, thus, we loaded all cells dissected from embryos without pre-counting, to minimize cell loss and maximize the number of captured single cells. Additional components used for library preparation include the Chromium Single Cell 3′ Library and Gel Bead Kit v2 (PN-120237) and the Chromium Single Cell 3′ Chip kit v2 (PN-120236).

Single cell library preparation and sequencing

Libraries were prepared according to the manufacturer's instructions using the Chromium Single Cell 3′ Library & Gel Bead Kit v2 (PN- 120237) and Chromium i7 Multiplex Kit (PN-120262). Final libraries were sequenced on the NextSeq 500 and Hiseq 4000. Somite-matched WT and *Hand2*-null replicate libraries from each litter were pooled and sequenced in the same lane (**Table 5.2**). Sequencing parameters were selected according to the Chromium Single Cell v2 specifications. All libraries were sequenced to a mean read depth of at least 50,000 total aligned reads/cell (**Table 5.3**).

Processing of sequencing data

Raw sequencing reads were processed using the Cell Ranger v2.2.0 pipeline from 10X Genomics. Briefly, reads were demultiplexed, aligned to the mouse mm10 genome

and UMI counts were quantified per gene per cell to generate a gene-barcode matrix (**Table 5.4**). Data from multiple samples (WT only analysis, WT/*Hand2*-null E7.75/E8.25 and WT/*Hand2*-null E9.25 analyses) were aggregated and normalized to the same sequencing depth, resulting in a combined gene-barcode matrix of all samples.

Cell filtering and cell-type clustering analysis

We sequenced the transcriptomes of 36,777 cells captured from WT and 37,149 cells captured from *Hand2*-null embryos in total (**Table 5.1**). Further filtering and clustering analyses of these cells were performed with the Seurat v2.2 R package, as described in the tutorials (http://satijalab.org/seurat/). For each aggregated dataset (WT only, WT/*Hand2*-null E7.75/E8.25, WT/*Hand2*-null E9.25), cells were normalized for genes expressed per cell and total expression, then multiplied by a scale factor of 10,000 and log-transformed. Cells that were of low quality or represented doublets were excluded from our analyses - this was achieved by filtering out cells with greater than 8000 and fewer than 1500 genes in Seurat (**Table 5.5**). We then performed a linear regression on all genes to eliminate technical variability due to the number of genes detected, embryonic time point, embryo replicate and stage of the cell cycle (*ScaleData* Function). For the *Hand2*-null analyses, we also regressed out the *Hand2* gene to eliminate its contribution to cell clustering. Highly variable genes in the dataset were computed and used as input for Principal Component Analysis. Significant PCs were used for downstream graph-based, semi-unsupervised clustering into distinct populations (*FindClusters* Function) and Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction was used to project these populations in 2D[203]. For

clustering, the resolution parameter, which indirectly controls the number of clusters, was approximated based on the number of cells according to Seurat guidelines; a vector of resolution parameters was passed to the *FindClusters* function and the optimal resolution that established discernible clusters with distinct marker gene expression was selected. One or two cell clusters would emerge that expressed marker genes representing multiple populations; these contained cells with low UMI and gene counts that escaped the first filtering step. These cells were removed from the analyses. To identify marker genes, the clusters were compared pairwise for differential gene expression using the Wilcoxon rank sum test for single-cell gene expression (*FindAllMarkers* function, min.pct = 0.25, min.diff.pct = 0.1, return.thresh (p-value cut-off) = $1x10^{-4}$). To assign identities to these subpopulations, we cross-referenced their marker genes with known cardiac subtype markers and *in situ* hybridization data from the literature (**Supplementary Table 2.1**). We also validated several of these marker genes by fluorescence in situ hybridization (**Fig. 2.4**). We removed blood, endoderm-, and ectoderm- derived clusters based on their expression of known blood markers such as the hemoglobin genes, endoderm markers such as *Epcam* and *Foxa2*, and ectoderm markers such as *Pou3f1* and *Sox2* (**Fig. 2.2c**), retaining the cells of mesodermal or neural crest identity. The clustering approach was then repeated for these retained mesodermal and neural crest cells, beginning with the regression of technical variables, identification of highly variable genes, Principal Component Analysis, graph-based clustering, UMAP projection and marker analysis. Similarly, all reclustering analyses (**Fig. 2.3**) were processed as described above. For computing differentially expressed genes between *Pitx2*-positive (normalized UMI >0.1) and *Pitx2*-negative (normalized

UMI < 0.1) cells in clusters A and I (**Fig. 2.9a**) and for the *Hand2*-null analyses, the *FindMarkers* function was used on WT and *Hand2*-null cells from each cluster with the following parameters: Wilcoxon rank sum test, min.diff.pct = 0.1, min.pct = 0.25, logFC.threshold = 0.2. Additionally, the max.cells.per.ident argument was used to ensure that equivalent numbers of cells were considered in the analysis. For each population analyzed, the number of cells (n) given to this argument was set to the population/genotype that had the lower cell number. An adjusted *p*-value (Bonferroni Correction) cut-off < $1x10^{-4}$ was used to identify differentially expressed genes.

Prediction of cell fate determinants

Cell fate determinants for OFT and RV from the AHF were predicted using a modified version of the method that we previously developed. This procedure was performed on E7.75 and E8.25 WT OFT, RV and AHF cells. One hundred cells were randomly selected from the AHF and from the OFT and RV daughter populations and the normalized ratio difference (NRD) was computed for all combinations of these 100 cells, yielding 10,000 parent-daughter cell combinations. The NRD was calculated for all pairs of differentially expressed TFs between OFT, RV and AHF cells and averaged over the 10,000 cell combinations. We minimized the impact of zero inflation by relying on the highly stringent adjusted (Bonferroni) p-value cutoff (< $1x10^{-4}$) and by filtering out very lowly expressed genes in defining differentially expressed TFs (average expression >= 0.5 normalized UMI). We filtered out genes that were detected in less than 25% of cells in populations being compared, and only tested genes that show a minimum difference of 0.1 in the fraction of detection between the two groups. TF pairs whose mean NRD

was more than 0.05 in one lineage direction but less than 0.01 in the other lineage direction were selected. Finally, the TF pairs that resided in the strongly connected component of the GRN were kept as the final candidate cell fate determinants.

Cell trajectory analysis

Pseudotime analyses were performed using the Monocle 2 package, as described in the tutorials (http://cole-trapnell-lab.github.io/monocle-release/). Differentially expressed genes, as determined in Seurat using the *FindAllMarkers* function, between the myocardium, CPC and WT/*Hand2*-null AHF, OFT and RV cells were used as input for temporal ordering of these cells along the differentiation trajectory.

***In situ* hybridization experiments**

Each *in situ* hybridization experiment was replicated at least twice for identifying spatial expression of genes and three times for quantification of *in situ* signal for differentially expressed genes in the *Hand2*-null analysis. For whole-mount experiments: de-yolked whole embryos were fixed in a 4% formaldehyde solution (ThermoFisher Scientific, CAT# 28906) overnight at 4°C followed by 2X PBST washes and 5-minute incubations in a dehydration series of 25%, 50%, 75% and 100% methanol (Fisher Scientific, CAT# A454-1). At this point embryos were stored in 100% methanol at -20°C until the *in situ* protocol was initiated. Yolk sac DNA was used for genotyping. The whole-mount *in situ* assay was adapted from the protocol formulated for whole-mount zebrafish embryos[204] using the RNAscope Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics, CAT# 323100), with minor modifications (the air-drying step was excluded

in our protocol, Protease Plus was used for embryo permeabilization, and the 0.2X SSCT wash step between reagent incubations was reduced to 3X 8 mins). Whole-mount embryos were imaged in 0.1% PBST using the Leica M165 fluorescent dissecting scope (FC/PLANAPO 1.0x #10450028; Camera – DFC3000G; ET GFP #10447408, ET mCHER #10450195, Acquisition Software – LAS V4.6). Quantification of transcript signal from whole-mount WT and *Hand2*-null embryos was performed using ImageJ v1.51m9. The mean grey value and integrated density of a defined area, that was kept consistent between WT and mutant embryos, as well as the background fluorescence level for the same defined area per embryo was measured. The corrected total fluorescence (CTF) for each gene was calculated using the following formula: Integrated density – (Area x Background fluorescence). CTF values were the $\log_{10}$ transformed before *t*-tests were conducted to satisfy prerequisite assumptions of normality.

For *in situ* hybridization experiments performed on embryo and postnatal day 1 (p1) heart sections: embryos and p1 hearts were washed 3X in PBS after overnight fixation in 4% formaldehyde and stored in 70% ethanol (VWR, CAT# 89125-186) indefinitely until embedding. Embryos were embedded in Histogel (Thermo Scientific, CAT# R904012) and paraffin processed using standard protocols and embedded for transverse sectioning, while p1 hearts were directly processed in paraffin. Tissue slices were serially sectioned at 5 μm intervals, mounted on slides and stored at room temperature until initiation of the RNAscope protocol for paraffin embedded sections (User manual catalog number 322452-USM). Sections were imaged with a Zeiss Axio Observer.Z1 inverted epifluorescence microscope (Carl Zeiss Microscopy, Thornwood, NY) with Zeiss Axiocam MRm and PCO.edge sCMOS (PCO.Imaging, Kelheim,

Germany) monochrome cameras run by Zeiss Zen imaging software. For 3D expression reconstruction, embryos were embedded in low melting agarose and imaged with a Zeiss lightsheet Z.1 selective plane illumination microscope (Carl Zeiss Microscopy) with 488nm and 561nm lasers, tandem PCO.edge sCMOS cameras and Zeiss Zen imaging software. 3D reconstructions of multi-view images were performed using Bitplane Imaris software v.9.0.2 (Andor Technology PLC).

Catalog numbers for RNAscope probes used: *Cck*, 402271-C3; *Cited1*, 432471; *Hand1*, 429651-C2; *Irx4*, 504831; *Pln*, 506241; *Rgs5*, 430181; *Sema3c*, 441441-C3; *Tdgf1*, 506411; *Tbx5*, 519581-C2 and *Upp1*, 504841-C2; *Tbx1,* 481911-C2; *Nr2f2,* 480301-C3; *Hoxb1, 541861*; *LacZ*, 313451; *Tbx18,* 515221-C2; *Mab21l2,* 456901; *Wnt5a,* 316791; *Bmp4,* 401301- C2; *3632451O06Rik,* 502031; *C130080G10Rik,* 506051; *Tnnt2,* 418681-C4; *Actc1,* 510361-C2; *Cacna2d2*, 449221-C2; *Tbx2*, 448991-C2; *Shox2*, 554291-C3; *TdTomato*, 317041-C2.

Statistics and reproducibility

Standard statistical analyses were performed using GraphPad Prism 8. The number of replicates for each experiment are described in the figure legends. All differentially expressed genes were calculated using the Wilcoxon rank sum test (two-sided) and a Bonferroni correction adjusted p-value < $1x10^{-4}$ was considered statistically significant. The level of significance in all graphs is represented as follow: **P*<0.05 and ***P*<0.01. For fluorescence quantification, corrected total fluorescence values were log transformed before *t*-tests were conducted to satisfy the prerequisite assumptions of normality. For all quantifications, the mean +/- s.e.m is reported. DAVID v6.8 was used

for GO term enrichment analysis; significant functional enrichment was statistically determined using a modified Fisher's exact test (EASE score) followed by Benjamini-Hochberg correction for multiple comparisons, with 0.01 as a p-value cut-off. No experimental samples were excluded from the statistical analyses. Sample size was not pre- determined through power calculations, and no randomization or investigator blinding approaches were implemented during the experiments and data analyses. When representative results are presented to indicate expression patterns of genes in WT embryos, at least two independent embryos were analyzed.

Data availability

All source data, including sequencing reads and single-cell expression matrices were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE126128. Data underlying each figure are available in the Supplementary Tables and on the UCSC cell browser at https://mouse-cardiac.cells.ucsc.edu. Users can use the cell browser to explore the data, view expression of genes of interest in each UMAP plot and download datasets for custom processing and analysis.

Table 3.1: Breakdown of cell proportions per genotype, per embryonic stage for each population discussed in the *Hand2*-null analysis.

Population	Embryos Stage	WT cells raw	*Hand2*$^{-/-}$ cells raw	WT cell %	*Hand2*$^{-/-}$ cell %
AHF	E7.75	637	1019	5.72	8.29129373
AHF	E8.25	410	458	4.22	4.07981472
AHF	E9.25	406	986	2.54	7.23245067
pSHF	E7.75	304	1248	2.73	10.1545972
pSHF	E8.25	681	1679	7.02	14.9563513
pSHF	E9.25	893	507	5.59	3.71891733
OFT	E7.75	253	354	2.27	2.88039056
OFT	E8.25	627	276	6.46	2.4585783
OFT	E9.25	515	362	3.22	2.65532165
RV	E7.75	161	132	1.44	1.07404394
RV	E8.25	418	331	4.31	2.94851238
RV	E9.25	227	369	1.42	2.70666764
LV	E7.75	239	201	2.14	1.635476
LV	E8.25	673	492	6.94	4.38268306
LV	E9.25	603	343	3.77	2.51595394
AVC	E7.75	32	86	0.28	0.6997559
AVC	E8.25	122	156	1.25	1.38963121
Atrial	E7.75	23	76	0.20	0.61838893
Atrial	E8.25	131	212	1.35	1.88847319
SV	E7.75	436	419	3.91	3.40927583
SV	E8.25	391	508	4.03	4.52520934

Table 5.1: Sample features in single-cell RNA sequencing experiments

Genotype and Embryo Replicate	Embryonic Stage	Sex	Total Cells (Post-Filtering)
WT 1	E7.75	F	3,161
WT 2	E7.75	F	1,710
WT 3	E7.75	F	2,761
WT 4	E7.75	M	1,326
WT 5	E7.75	M	2,289
Hand2-KO 1	E7.75	F	5,221
Hand2-KO 2	E7.75	F	4,228
Hand2-KO 3	E7.75	F	2,887
WT 1	E8.25	M	2,389
WT 2	E8.25	F	7,465
Hand2-KO 1	E8.25	M	6,097
Hand2-KO 2	E8.25	M	5,173
WT 1	E9.25	F	4,761
WT 2	E9.25	M	11, 361
Hand2-KO 1	E9.25	F	8,510
Hand2-KO 2	E9.25	M	5,525

Table 5.2: Summary of sequencing metrics per sample library

Embryonic Stage and Genotype	Q30 bases in barcode (%)	Q30 bases in RNA read (%)	Q30 Score in sample index (%)	Q30 Score in UMI (%)	Sequencing saturation (%)
E7.75 WT	97.6	68.6	93.1	97.6	55.9
E7.75 WT	96.6	69.5	91	95.8	43.8
E7.75 WT	96.6	69.9	90.7	95.7	34.3
E7.75 WT	96.6	69.1	91.1	95.7	49
E7.75 WT	96.6	69.3	90.8	95.7	43.6
E7.75 *Hand2*-KO	97.5	65.8	92.3	97.5	49.3
E7.75 *Hand2*-KO	95.3	74.4	93.4	94.2	40.4
E7.75 *Hand2*-KO	94.3	62.8	89.5	93.8	41.2
E8.25 WT	96.5	68.8	89.4	96.5	69.5
E8.25 WT	97.3	71.3	91.3	97.2	40.5
E8.25 *Hand2*-KO	97.3	72.8	91.5	97.1	43.5
E8.25 *Hand2*-KO	97.3	73.6	91.3	97.2	51.9
E9.25 WT	95.4	77.2	93.4	94.3	41.7
E9.25 WT	95.1	66.1	88.4	94.8	48.8
E9.25 *Hand2*-KO	95.2	66	88.6	94.9	59.9
E9.25 *Hand2*-KO	95.1	75.8	93.1	94.1	50.9

Table 5.3: Median reads, genes and UMI counts per cell per sample

Embryonic Stage and Genotype	Pre-Normalization Median Reads/Cell	Median Genes/Cell	Median UMI Counts/Cell
E7.75 WT	119,043	5,375	28,820
E7.75 WT	111,758	5,419	29,188
E7.75 WT	69,637	4,814	22,156
E7.75 WT	141,134	5,760	33,421
E7.75 WT	91,262	4,957	23,645
E7.75 *Hand2*-KO	90,424	4,888	22,951
E7.75 *Hand2*-KO	60,823	4,523	19,027
E7.75 *Hand2*-KO	72,174	4,798	20,666
E8.25 WT	155,229	5,237	24,751
E8.25 WT	55,971	4,357	17,483
E8.25 *Hand2*-KO	69,266	4,573	19,583
E8.25 *Hand2*-KO	82,687	4,520	19,026
E9.25 WT	58,167	4,362	17,245
E9.25 WT	64, 432	4,182	15,770
E9.25 *Hand2*-KO	85,275	4,134	15,624
E9.25 *Hand2*-KO	60,553	3,924	14,621

Table 5.4: Percentage of reads mapped to transcriptome features per sample

Embryonic Stage and Genotype	Transcriptome (%)	Exonic regions (%)	Intronic regions (%)	Intergenic regions (%)	Antisense to gene (%)
E7.75 WT	62.6	65	9.6	4.3	3
E7.75 WT	61.3	63.4	7.3	3.3	3.5
E7.75 WT	60.2	62.3	7.8	3.9	3.4
E7.75 WT	61.1	63.4	7.6	3.4	3.5
E7.75 WT	60.1	62.3	8.1	3.6	3.5
E7.75 *Hand2*-KO	61.6	63.9	8.8	3.7	3
E7.75 *Hand2*-KO	60.4	62.8	9.9	4.1	3.6
E7.75 *Hand2*-KO	56.6	58.9	8.6	3.6	3.6
E8.25 WT	63.4	65.6	10	3.1	3
E8.25 WT	63.1	65.5	10.1	4.3	3.1
E8.25 *Hand2*-KO	62.7	65.1	10.9	4.1	3
E8.25 *Hand2*-KO	63.8	66.2	11.2	3.7	3
E9.25 WT	63.2	65.7	10.7	3.2	3.4
E9.25 WT	60.4	62.5	8.9	3.1	3.1
E9.25 *Hand2*-KO	58.3	60.4	11.1	3.2	3.1
E9.25 *Hand2*-KO	63.2	65.4	9.6	3.2	2.9

Table 5.5: Number of cells in each germ lineage per sample after quality control

Embryonic Stage and Genotype	# Endoderm Cells	# Ectoderm Cells	# Mesoderm Cells	Total Cells (Post-Filtering)
E7.75 WT	874	1,257	946	3,077
E7.75 WT	595	403	697	1,695
E7.75 WT	752	928	1,067	2,747
E7.75 WT	627	69	624	1,320
E7.75 WT	804	384	1,097	2,285
E7.75 *Hand2*-KO	1,108	1,404	2,688	5,200
E7.75 *Hand2*-KO	603	1,486	2,130	4,219
E7.75 *Hand2*-KO	610	627	1,634	2,871
E8.25 WT	345	2	1,973	2,320
E8.25 WT	1,432	2,252	3,692	7,376
E8.25 *Hand2*-KO	964	1,655	3,458	6,077
E8.25 *Hand2*-KO	745	934	3,470	5,149
E9.25 WT	964	221	3,482	4,667
E9.25 WT	2,686	2,492	6,112	11,290
E9.25 *Hand2*-KO	1,796	695	5,862	8,353
E9.25 *Hand2*-KO	1,402	484	3,394	5,280

Supplementary Table 2.1 -
https://www.dropbox.com/s/vidihu63froin0u/Supplementary%20Table%202.1.xlsx?dl=0

Supplementary Table 3.1 -
https://www.dropbox.com/s/m9hnc5yu6twf8e3/Supplementary%20Table%203.1.xlsx?dl=0

References

1. Teun van der Bom, A. Carla Zomer, Aeilko H. Zwinderman, Folkert J. Meijboom, B. J. B. and B. J. M. M. The changing epidemiology of congenital heart disease. *Nat. Rev. Cardiol.* **8**, 50–60 (2011).
2. Zaidi, S. & Brueckner, M. Genetics and Genomics of Congenital Heart Disease. *Circ. Res.* **120**, 923–940 (2017).
3. Fahed, A. C., Gelb, B. D., Seidman, J. G. & Seidman, C. E. Genetics of Congenital Heart Disease. *Circ. Res.* **112**, 707–720 (2013).
4. Liu, Y. *et al.* Global birth prevalence of congenital heart defects 1970 – 2017 : updated systematic review and meta-analysis of 260 studies. *Int. J. Epidemiol.* 455–463 (2019). doi:10.1093/ije/dyz009
5. Jin, S. C. *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601 (2017).
6. Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126**, 1037–48 (2006).
7. Schott, J.J. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–111 (1998).
8. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443–447 (2003).
9. Kodo, K. *et al.* Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects. *Circ. J.* **76**, 1703–1711 (2012).
10. Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).

11. Tsuchihashi, T. *et al.* Hand2 function in second heart field progenitors is essential for cardiogenesis. *Dev. Biol.* **351**, 62–69 (2011).

12. Zaffran, S. & Frasch, M. Early signals in cardiac development. *Circ. Res.* **91**, 457–469 (2002).

13. Devine, W. P., Wythe, J. D., George, M., Koshiba-Takeuchi, K. & Bruneau, B. G. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *Elife* **3**, 1–23 (2014).

14. Lescroart, F. *et al.* Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat. Cell Biol.* **16**, 829–840 (2014).

15. Bondue, A. *et al.* Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell* **3**, 69–84 (2008).

16. Meilhac, S. M. & Buckingham, M. E. The deployment of cell lineages that form the mammalian heart. *Nat. Rev. Cardiol.* **15**, 705–724 (2018).

17. Andersen, P. *et al.* Precardiac organoids form two heart fields via Bmp/Wnt signaling. *Nat. Commun.* **9**, (2018).

18. Cai, C.L. *et al.* Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* **5**, 877–89 (2003).

19. Chen, L., Fulcoli, F. G., Tang, S. & Baldini, A. Tbx1 regulates proliferation and differentiation of multipotent heart progenitors. *Circ. Res.* **105**, 842–851 (2009).

20. Ku, M., Gessert, S. & Ku, M. The Multiple Phases and Faces of Wnt Signaling During Cardiac Differentiation and Development Wnt Signaling Pathways : A Short

Synopsis. 1798–1806 (2010). doi:10.1161/CIRCRESAHA.110.221531

21. Kelly, R. G., Buckingham, M. E. & Moorman, A. F. Heart fields and cardiac morphogenesis. *Cold Spring Harb. Perspect. Med.* **4**, 1–11 (2014).
22. Pucéat, M. Embryological origin of the endocardium and derived valve progenitor cells: From developmental biology to stem cell-based valve repair. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 917–922 (2013).
23. Desgrange, A., Le Garrec, J.F. & Meilhac, S. M. Left-right asymmetry in heart development and disease: forming the right loop. *Development* **145**, (2018).
24. Epstein, J. A., Aghajanian, H. & Singh, M. K. Semaphorin signaling in cardiovascular development. *Cell Metab.* **21**, 163–173 (2015).
25. Miyagawa-tomita, S., Arima, Y. & Kurihara, H. The “Cardiac Neural Crest” Concept Revisited. 3–8 doi:10.1007/978-4-431-54628-3
26. Plein, A., Fantin, A. & Ruhrberg, C. *Neural Crest Cells in Cardiovascular Development*. *Neural Crest and Placodes* **111**, (Elsevier Inc., 2015).
27. Combs, M. D. & Yutzey, K. E. Heart valve development: Regulatory networks in development and disease. *Circ. Res.* **105**, 408–421 (2009).
28. Meilhac, S. M., Lescroart, F., Blanpain, C. D. & Buckingham, M. E. Cardiac cell lineages that form the heart. *Cold Spring Harb. Perspect. Med.* **4**, (2014).
29. Braitsch, C. M., Combs, M. D., Quaggin, S. E. & Yutzey, K. E. Pod1/Tcf21 is regulated by retinoic acid signaling and inhibits differentiation of epicardium-derived cells into smooth muscle in the developing heart. *Dev. Biol.* **368**, 345–357 (2012).
30. del Monte-Nieto, G. *et al.* Control of cardiac jelly dynamics by NOTCH1 and NRG1 defines the building plan for trabeculation. *Nature* **557**, 439–445 (2018).

31. van Eif, V. W. W. *et al.* Transcriptome analysis of mouse and human sinoatrial node cells reveals a conserved genetic program. *Development* **146**, (2019).

32. Saxon, J. G. *et al.* BMP2 expression in the endocardial lineage is required for AV endocardial cushion maturation and remodeling. *Dev. Biol.* **430**, 113–128 (2017).

33. Katano, W., Moriyama, Y., Takeuchi, J. K. & Koshiba-Takeuchi, K. Cardiac septation in heart development and evolution. *Dev. Growth Differ.* **61**, 114–123 (2019).

34. Hirokawa, N., Tanaka, Y., Okada, Y. & Takeda, S. Nodal Flow and the Generation of Left-Right Asymmetry. *Cell* **125**, 33–45 (2006).

35. Francis, R. J. B., Christopher, A., Devine, W. A., Ostrowski, L. & Lo, C. Congenital heart disease and the specification of left-right asymmetry. *Am. J. Physiol. Circ. Physiol.* **302**, H2102–H2111 (2012).

36. Dykes, I. Left Right Patterning, Evolution and Cardiac Development. *J. Cardiovasc. Dev. Dis.* **1**, 52–72 (2014).

37. Yoshioka, H. *et al.* Pitx2, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry. *Cell* **94**, 299–305 (1998).

38. Tsukui, T. *et al.* Multiple left-right asymmetry defects in Shh-/- mutant mice unveil a convergence of the Shh and retinoic acid pathways in the control of Lefty-1. *Proc. Natl. Acad. Sci.* **96**, 11376–11381 (2002).

39. Campione, M. *et al.* Pitx2 Expression Defines a Left Cardiac Lineage of Cells: Evidence for Atrial and Ventricular Molecular Isomerism in the iv/iv Mice. *Dev. Biol.* **231**, 252–264 (2001).

40. Mercola, M. Left-right asymmetry: Nodal points. *J. Cell Sci.* **116**, 3251–3257 (2003).

41. Franco, D., Christoffels, V. M. & Campione, M. Homeobox transcription factor Pitx2: The rise of an asymmetry gene in cardiogenesis and arrhythmogenesis. *Trends Cardiovasc. Med.* **24**, 23–31 (2014).

42. Ai, D. *et al.* Pitx2 regulates cardiac left–right asymmetry by patterning second cardiac lineage-derived myocardium. *Dev. Biol.* **296**, 437–449 (2006).

43. Bajolle, F. *et al.* Rotation of the Myocardial Wall of the Outflow Tract Is Implicated in the Normal Positioning of the Great Arteries. *Circ. Res.* **98**, 421–428 (2006).

44. Galli, D. *et al.* Atrial myocardium derives from the posterior region of the second heart field, which acquires left-right identity as Pitx2c is expressed. *Development* **135**, 1157–1167 (2008).

45. Franco, D., Sedmera, D. & Lozano-Velasco, E. Multiple Roles of Pitx2 in Cardiac Development and Disease. *J. Cardiovasc. Dev. Dis.* **4**, 16 (2017).

46. Francou, A. *et al.* Second heart field cardiac progenitor cells in the early mouse embryo. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 795–798 (2013).

47. Zaffran, S., Robrini, N. & Bertrand, N. Retinoids and Cardiac Development. *J. Dev. Biol.* **2**, 50–71 (2014).

48. Stefanovic, S. & Zaffran, S. Mechanisms of retinoic acid signaling during cardiogenesis. *Mech. Dev.* **143**, 9–19 (2017).

49. Duester, G. Retinoic Acid Synthesis and Signaling during Early Organogenesis. *Cell* **134**, 921–931 (2008).

50. Ryckebusch, L. *et al.* Retinoic acid deficiency alters second heart field formation. **105**, (2008).

51. Xavier-Neto, J. *et al.* A retinoic acid-inducible transgenic marker of sino-atrial development in the mouse heart. *Development* **126**, 2677–87 (1999).

52. Vincent, D. & Buckingham, M. E. How to Make a Heart : The Origin and Regulation of Cardiac Progenitor Cells. **90**, 1–41 (2010).

53. Niederreither, K. *et al.* Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. *Development* **128**, 1019–31 (2001).

54. Roux, M. & Zaffran, S. Hox Genes in Cardiovascular Development and Diseases. (2016). doi:10.3390/jdb4020014

55. Bertrand, N. *et al.* Hox genes de fi ne distinct progenitor sub-domains within the second heart fi eld. **353**, 266–274 (2011).

56. Rana, M. S. *et al.* Tbx1 coordinates addition of posterior second heart field progenitor cells to the arterial and venous poles of the heart. *Circ. Res.* **115**, 790–799 (2014).

57. Bono, C. De *et al.* T-box genes and retinoic acid signaling regulate the segregation of arterial and venous pole progenitor cells in the murine second heart field. *Human Molecular Genetics* **27**, 3747–3760 (2018).

58. Xie, L. *et al.* Tbx5-Hedgehog Molecular Networks Are Essential in the Second Heart Field for Atrial Septation. *Dev. Cell* **23**, 280–291 (2012).

59. Li, G. *et al.* Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. *Dev. Cell* **39**, 491–507 (2016).

60. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).

61. Klein, A. M. & Treutlein, B. Single cell analyses of development in the modern era.

Development **146**, dev181396 (2019).

62. Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**, 1308-1323.e30 (2016).
63. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
64. Dahlin, J. S. *et al.* A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood* **131**, e1–e11 (2018).
65. Stévant, I. *et al.* Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell Rep.* **22**, 1589–1599 (2018).
66. Olsson, A. *et al.* Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537**, 698–702 (2016).
67. Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
68. DeLaughter, D. M. *et al.* Single-Cell Resolution of Temporal Gene Expression during Heart Development. *Dev. Cell* **39**, 480–490 (2016).
69. Lescroart, F. *et al.* Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* **4174**, 1–9 (2018).
70. Jia, G. *et al.* Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat. Commun.* **9**, (2018).
71. Sahara, M. *et al.* Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract. *Dev. Cell* **48**, 475-490.e7 (2019).
72. Salomon, R. *et al.* Droplet-based single cell RNAseq tools: a practical guide. *Lab*

Chip **19**, 1706–1727 (2019).

73. Srivastava, D., Cserjesi, P. & Olson, E. N. A Subclass of bHLH Proteins Required for Cardiac Morphogenesis. *Science* **270**, 1995–1999 (1995).

74. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHand. *Nat. Genet.* **16**, 154–160 (1997).

75. McFadden, D. G. The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189–201 (2004).

76. Barbosa, A. C. *et al.* Hand transcription factors cooperatively regulate development of the distal midline mesenchyme. *Dev. Biol.* **310**, 154–168 (2007).

77. George, R. M. & Firulli, A. B. Hand Factors in Cardiac Development. *Anat. Rec.* **107**, 101–107 (2018).

78. Thomas, T., Yamagishi, H., Overbeek, P. A., Olson, E. N. & Srivastava, D. The bHLH factors, dHAND and eHAND, specify pulmonary and systemic cardiac ventricles independent of left-right sidedness. *Dev. Biol.* **196**, 228–36 (1998).

79. Biben, C. & Harvey, R. P. Homeodomain factor Nkx2-5 controls left/right asymmetric expression of bHLH gene eHand during murine heart development. *Genes Dev.* **11**, 1357–1369 (1997).

80. Morikawa, Y. & Cserjesi, P. Cardiac Neural Crest Expression of Hand2 Regulates Outflow and Second Heart Field Development. *Circ. Res.* **103**, 1422–1429 (2008).

81. Holler, K. L. *et al.* Targeted deletion of Hand2 in cardiac neural crest-derived cells influences cardiac gene expression and outflow tract development. *Dev. Biol.* **341**,

291–304 (2010).

82. Hendershot, T. J. *et al.* Conditional deletion of Hand2 reveals critical functions in neurogenesis and cell type-specific gene expression for development of neural crest-derived noradrenergic sympathetic ganglion neurons. *Dev. Biol.* **319**, 179–191 (2008).

83. Galli, A. *et al.* Distinct Roles of Hand2 in Initiating Polarity and Posterior Shh Expression during the Onset of Mouse Limb Bud Development. *PLoS Genet.* **6**, e1000901 (2010).

84. Barnes, R. M. *et al.* Hand2 loss-of-function in hand1-expressing cells reveals distinct roles in epicardial and coronary vessel development. *Circ. Res.* **108**, 940–949 (2011).

85. Laurent, F. *et al.* HAND2 Target Gene Regulatory Networks Control Atrioventricular Canal and Cardiac Valve Development. *Cell Rep.* **19**, 1602–1613 (2017).

86. McCulley, D. J. & Black, B. L. Transcription Factor Pathways and Congenital Heart Disease. *Current Topics in Developmental Biology* **100**, (Elsevier Inc., 2012).

87. Nam, Y.-J. *et al.* Reprogramming of human fibroblasts toward a cardiac fate. *Proc. Natl. Acad. Sci.* **110**, 5588–5593 (2013).

88. Harris, I. S. & Black, B. L. Development of the endocardium. *Pediatr. Cardiol.* **31**, 391–399 (2010).

89. Takase, H. *et al.* Genome-wide identification of endothelial cell-enriched genes in the mouse embryo. *Blood* **120**, 914–923 (2012).

90. Brachtendorf, G. *et al.* Early expression of endomucin on endothelium of the mouse embryo and on putative hematopoietic clusters in the dorsal aorta. *Dev. Dyn.* **222**,

410–419 (2001).

91. Haenig, Â., Kispert, A. & Kraus, F. Cloning and expression analysis of the mouse T-box gene Tbx18. *Mech. Dev.* **100**, 87–91 (2001).

92. von Gise, A. *et al.* WT1 regulates epicardial epithelial to mesenchymal transition through β-catenin and retinoic acid signaling pathways. *Dev. Biol.* **356**, 421–431 (2011).

93. Bhatt, S., Diaz, R. & Trainor, P. A. Signals and Switches in Mammalian Neural Crest Cell Differentiation. *Cold Spring Harb. Perspect. Biol.* **5**, a008326–a008326 (2013).

94. Sugii, H. *et al.* The Dlx5-FGF10 signaling cascade controls cranial neural crest and myoblast interaction during oropharyngeal patterning and development. *Development* **144**, 4037–4045 (2017).

95. Burgess, R., Cserjesi, P., Ligon, K. L. & Olson, E. N. Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. *Dev. Biol.* **168**, 296–306 (1995).

96. Magli, A., Schnettler, E., Rinaldi, F., Bremer, P. & Perlingeiro, R. C. R. Functional Dissection of Pax3 in Paraxial Mesoderm Development and Myogenesis. *Stem Cells* **31**, 59–70 (2013).

97. Lowe, L. A., Yamada, S. & Kuehn, M. R. HoxB6-Cre transgenic mice express Cre recombinase in extra-embryonic mesoderm, in lateral plate and limb mesoderm and at the midbrain/hindbrain junction. *Genesis* **26**, 118–120 (2000).

98. Mahlapuu, M., Ormestad, M., Enerback, S. & Carlsson, P. The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. *Development* **128**, 155–166 (2001).

99. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* (2019). doi:10.1038/s41586-019-0933-9

100. Lescroart, F. *et al.* Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. *Development* **137**, 3269–3279 (2010).

101. Lescroart, F. *et al.* Clonal analysis reveals a common origin between nonsomite-derived neck muscles and heart myocardium. *Proc. Natl. Acad. Sci.* **112**, 1446–1451 (2015).

102. Bao, Z. Regulation of Chamber-Specific Gene Expression in the Developing Heart by Irx4. *Science* **283**, 1161–1164 (1999).

103. Nelson, D. O., Jin, D. X., Downs, K. M., Kamp, T. J. & Lyons, G. E. Irx4 identifies a chamber-specific cell population that contributes to ventricular myocardium development. *Dev. Dyn.* **243**, 381–392 (2014).

104. Zhang, Z. *et al.* Functional Roles of Ca v 1.3(α 1D) Calcium Channels in Atria. *Circulation* **112**, 1936–1944 (2005).

105. Devalla, H. D. *et al.* Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. *EMBO Mol. Med.* **7**, 394–410 (2015).

106. Espinoza-Lewis, R. A. *et al.* Shox2 is essential for the differentiation of cardiac pacemaker cells by repressing Nkx2-5. *Dev. Biol.* **327**, 376–385 (2009).

107. Cambier, L., Plate, M., Sucov, H. M. & Pashmforoush, M. Nkx2-5 regulates cardiac growth through modulation of Wnt signaling by R-spondin3. *Development* **141**, 2959–2971 (2014).

108. Cai, X. *et al.* Myocardial Tbx20 regulates early atrioventricular canal formation and endocardial epithelial–mesenchymal transition via Bmp2. *Dev. Biol.* **360**, 381–390 (2011).

109. Barnes, R. M. *et al.* MEF2C regulates outflow tract alignment and transcriptional control of Tdgf1. *Development* **143**, 774–779 (2016).

110. Lee, D. *et al.* Endogenous transmembrane protein UT2 inhibits pSTAT3 and suppresses hematological malignancy. *J. Clin. Invest.* **126**, 1300–1310 (2016).

111. Ivey, K. *et al.* Gαq and Gα11 proteins mediate endothelin-1 signaling in neural crest-derived pharyngeal arch mesenchyme. *Dev. Biol.* **255**, 230–237 (2003).

112. Ivey, K. N. *et al.* Transcriptional Regulation During Development of the Ductus Arteriosus. *Circ. Res.* **103**, 388–395 (2008).

113. Offermanns, S. *et al.* Embryonic cardiomyocyte hypoplasia and craniofacial defects in Gα(q)/Gα11-mutant mice. *EMBO J.* **17**, 4304–4312 (1998).

114. Moon, A. M. *et al.* Disruption of G-Protein γ5 Subtype Causes Embryonic Lethality in Mice. *PLoS One* **9**, e90970 (2014).

115. Bruneau, B. G. *et al.* Chamber-specific cardiac expression of Tbx5 and heart defects in Holt- Oram syndrome. *Dev. Biol.* **211**, 100–108 (1999).

116. Dunwoodie, S. L., Rodriguez, T. a & Beddington, R. S. Msg1 and Mrg1, founding members of a gene family, show distinct patterns of gene expression during mouse embryogenesis. *Mech. Dev.* **72**, 27–40 (1998).

117. Ilagan, R. *et al.* Fgf8 is required for anterior heart field development. *Development* **133**, 2435–2445 (2006).

118. Shaikh, S. A., Sahoo, S. K. & Periasamy, M. Phospholamban and sarcolipin: Are

they functionally redundant or distinct regulators of the Sarco(Endo)Plasmic Reticulum Calcium ATPase? *J. Mol. Cell. Cardiol.* **91**, 81–91 (2016).

119. Lay, J. M., Gillespie, P. J. & Samuelson, L. C. Murine prenatal expression of cholecystokinin in neural crest, enteric neurons, and enteroendocrine cells. *Dev. Dyn.* **216**, 190–200 (1999).
120. Kim, K. H., Nakaoka, Y., Augustin, H. G. & Koh, G. Y. Myocardial Angiopoietin-1 Controls Atrial Chamber Morphogenesis by Spatiotemporal Degradation of Cardiac Jelly. *Cell Rep.* **23**, 2455–2466 (2018).
121. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
122. Houweling, A. C., Van Borren, M. M., Moorman, A. F. M. & Christoffels, V. M. Expression and regulation of the atrial natriuretic factor encoding gene Nppa during development and disease. *Cardiovasc. Res.* **67**, 583–593 (2005).
123. Saito, Y., Kojima, T. & Takahashi, N. Mab21l2 is essential for embryonic heart and liver development. *PLoS One* **7**, (2012).
124. Persaud, S. D. *et al.* All trans-retinoic acid analogs promote cancer cell apoptosis through non-genomic Crabp1 mediating ERK1/2 phosphorylation. *Sci. Rep.* **6**, 2–3 (2016).
125. Nascone, N. & Mercola, M. An inductive role for the endoderm in Xenopus cardiogenesis. *Development* **121**, 515–23 (1995).
126. Yuasa, S. *et al.* Transient inhibition of BMP signaling by Noggin induces cardiomyocyte differentiation of mouse embryonic stem cells. *Nat. Biotechnol.* **23**, 607–611 (2005).

127. Costello, I. *et al.* Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes Dev.* **29**, 2108–2122 (2015).

128. Madabhushi, M. & Lacy, E. Anterior Visceral Endoderm Directs Ventral Morphogenesis and Placement of Head and Heart via BMP2 Expression. *Dev. Cell* **21**, 907–919 (2011).

129. Cohen, E. D., Miller, M. F., Wang, Z., Moon, R. T. & Morrisey, E. E. Wnt5a and Wnt11 are essential for second heart field progenitor development. *Development* **139**, 1931–1940 (2012).

130. Okawa, S., Nicklas, S., Zickenrott, S., Schwamborn, J. C. & del Sol, A. A Generalized Gene-Regulatory Network Model of Stem Cell Differentiation for Predicting Lineage Specifiers. *Stem Cell Reports* **7**, 307–315 (2016).

131. Okawa, S. & del Sol, A. A computational strategy for predicting lineage specifiers in stem cell subpopulations. *Stem Cell Res.* **15**, 427–434 (2015).

132. Uribe, V., Bad, C. & Sanz-ezquerro, J. J. Arid3b is essential for second heart field cell deployment and heart patterning. 4168–4181 (2014). doi:10.1242/dev.109918

133. Jiang, L., Chen, H., Pinello, L. & Yuan, G.-C. GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol.* **17**, 144 (2016).

134. Jindal, A., Gupta, P., Jayadeva & Sengupta, D. Discovery of rare cells from voluminous single cell expression data. *Nat. Commun.* **9**, 4719 (2018).

135. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

136. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-

wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

137. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
138. Lacourse, K. A. *et al.* Pancreatic function in CCK-deficient mice: adaptation to dietary protein does not require CCK. *Am. J. Physiol. Liver Physiol.* **276**, G1302–G1309 (1999).
139. Goetze, J. P. *et al.* Cardiomyocyte Expression and Cell-specific Processing of Procholecystokinin. *J. Biol. Chem.* **290**, 6837–6843 (2015).
140. Gottlieb, P. D. *et al.* Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat. Genet.* **31**, 25–32 (2002).
141. Bamford, R. N. *et al.* Loss-of-function mutations in the EGF-CFC gene CFC1 are associated with human left-right laterality defects. *Nat. Genet.* **26**, 365–9 (2000).
142. Kosaki, K. *et al.* Characterization and Mutation Analysis of Human LEFTY A and LEFTY B, Homologues of Murine Genes Implicated in Left-Right Axis Development. *Am. J. Hum. Genet.* **64**, 712–721 (1999).
143. Schindler, R. F., Poon, K. L., Simrick, S. & Brand, T. The Popeye domain containing genes: essential elements in heart rate control. *Cardiovasc. Diagn. Ther.* **2**, 308–19 (2012).
144. Villanueva, M. P. *et al.* Genetic and Comparative Mapping of Genes Dysregulated in Mouse Hearts Lacking the Hand2 Transcription Factor Gene. *Genomics* **80**, 593–600 (2002).
145. Zhang, W., Vreeland, A. C. & Noy, N. The RNA-binding protein HuR regulates

protein nuclear import. *J. Cell Sci.* **1**, jcs.192096 (2016).

146. Thattaliyath, B. D., Firulli, B. A. & Firulli, A. B. The basic-helix-loop-helix transcription factor HAND2 directly regulates transcription of the Atrial Naturetic Peptide gene. *J. Mol. Cell. Cardiol.* **34**, 1335–1344 (2002).
147. Chen, L., Fulcoli, F. G., Tang, S. & Baldini, A. Tbx1 regulates proliferation and differentiation of multipotent heart progenitors. *Circ Res* **105**, 842–851 (2009).
148. Yu, F. *et al.* Knockdown of interferon-induced transmembrane protein 1 (IFITM1) inhibits proliferation, migration, and invasion of glioma cells. *J. Neurooncol.* **103**, 187–195 (2011).
149. Molyneaux, K. A. The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. *Development* **130**, 4279–4286 (2003).
150. King, S. J. *et al.* Nesprin-1 and nesprin-2 regulate endothelial cell shape and migration. *Cytoskeleton* **71**, 423–434 (2014).
151. Ugorski, M., Dziegiel, P. & Suchanski, J. Podoplanin - a small glycoprotein with many faces. *Am. J. Cancer Res.* **6**, 370–86 (2016).
152. Sinha, T. *et al.* Loss of Wnt5a disrupts second heart field cell deployment and may contribute to OFT malformations in DiGeorge syndrome. *Hum. Mol. Genet.* **24**, 1704–1716 (2015).
153. Dupays, L., Kotecha, S., Angst, B. & Mohun, T. J. Tbx2 misexpression impairs deployment of second heart field derived progenitor cells to the arterial pole of the embryonic heart. *Dev. Biol.* **333**, 121–131 (2009).
154. Togi, K. *et al.* Essential role of Hand2 in interventricular septum formation and trabeculation during cardiac development. *Biochem. Biophys. Res. Commun.* **343**,

144–151 (2006).

155. Lieberman, Y., Rokach, L. & Shay, T. CaSTLe – Classification of single cells by transfer learning: Harnessing the power of publicly available single cell RNA sequencing experiments to annotate new experiments. *PLoS One* **13**, e0205499 (2018).

156. Stein-O'Brien, G. L. *et al.* Decomposing Cell Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species. *Cell Syst.* **8**, 395-411.e8 (2019).

157. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).

158. Taylor, R. C., Cullen, S. P. & Martin, S. J. Apoptosis: controlled demolition at the cellular level. *Nat. Rev. Mol. Cell Biol.* **9**, 231–241 (2008).

159. D'Aniello, E. & Waxman, J. S. Input overload: Contributions of retinoic acid signaling feedback mechanisms to heart development and teratogenesis. *Dev. Dyn.* **244**, 513–523 (2015).

160. Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat. Rev. Mol. Cell Biol.* **16**, 110–123 (2015).

161. Bardot, E. *et al.* Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential. *Nat. Commun.* **8**, 14428 (2017).

162. Kochhar, D. M., Jiang, H., Penner, J. D., Johnson, A. T. & Chandraratna, R. A. The use of a retinoid receptor antagonist in a new model to study vitamin A-dependent developmental events. *Int. J. Dev. Biol.* **42**, 601–8 (1998).

163. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of

regulatory variation. *Nature* **523**, 486–490 (2015).

164. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
165. Altemose, N., Maslan, A., Lai, A., White, J. A. & Streets, A. M. μDamID: a microfluidic approach for imaging and sequencing protein-DNA interactions in single cells. *BioRxiv* (2019). doi:http://dx.doi.org/10.1101/706903
166. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
167. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).
168. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
169. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).
170. Blow, M. J. *et al.* ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810 (2010).
171. McFadden, D. G. *et al.* A GATA-dependent right ventricular enhancer controls dHAND transcription in the developing heart. *Development* **127**, 5331–41 (2000).
172. Verzi, M. P., McCulley, D. J., De Val, S., Dodou, E. & Black, B. L. The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev. Biol.* **287**, 134–145 (2005).
173. Kester, L. & van Oudenaarden, A. Single-Cell Transcriptomics Meets Lineage

Tracing. *Cell Stem Cell* **23**, 166–179 (2018).

174. Kalhor, R. *et al.* Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
175. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
176. McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
177. Junker, J. P. *et al.* Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars. *bioRxiv* (2017). doi:10.1101/056499
178. Chabab, S. *et al.* Uncovering the Number and Clonal Dynamics of Mesp1 Progenitors during Heart Morphogenesis. *Cell Rep.* **14**, 1–10 (2016).
179. Skelly, D. A. *et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Rep.* **22**, 600–610 (2018).
180. Xiong, H. *et al.* Single-Cell Transcriptomics Reveals Chemotaxis Mediated Intra-Organ Crosstalk During Cardiogenesis. *Circ. Res.* (2019). doi:10.1161/CIRCRESAHA.119.315243
181. Kalamakis, G. *et al.* Quiescence Modulates Stem Cell Maintenance and Regenerative Capacity in the Aging Brain. *Cell* **176**, 1407-1419.e14 (2019).
182. Zaffaroni, G., Okawa, S., Morales-Ruiz, M. & del Sol, A. An integrative method to predict signalling perturbations for cellular transitions. *Nucleic Acids Res.* **47**, e72–e72 (2019).
183. Kumar, M. P. *et al.* Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **25**, 1458-

1468.e4 (2018).

184. Pérez-Pomares, J. M. & de la Pompa, J. L. Signaling During Epicardium and Coronary Vessel Development. *Circ. Res.* **109**, 1429–1442 (2011).

185. Kodo, K. *et al.* Regulation of Sema3c and the Interaction between Cardiac Neural Crest and Second Heart Field during Outflow Tract Development. *Sci. Rep.* **7**, 6771 (2017).

186. Hoffmann, A. D., Peterson, M. A., Friedland-Little, J. M., Anderson, S. A. & Moskowitz, I. P. sonic hedgehog is required in pulmonary endoderm for atrial septation. *Development* **136**, 1761–1770 (2009).

187. Okawa, S. *et al.* Transcriptional synergy as an emergent property defining cell subpopulation identity enables population shift. *Nat. Commun.* **9**, 2595 (2018).

188. Yabrodi, M. & Mastropietro, C. W. Hypoplastic left heart syndrome: from comfort care to long-term survival. *Pediatr. Res.* **81**, 142–149 (2017).

189. Shen, L. *et al.* Transcription factor HAND2 mutations in sporadic Chinese patients with congenital heart disease. *Chin. Med. J. (Engl).* **123**, 1623–7 (2010).

190. Sun, Y-M. *et al.* A HAND2 Loss-of-Function Mutation Causes Familial Ventricular Septal Defect and Pulmonary Stenosis. *Genes*|*Genomes*|*Genetics* **6**, 987–992 (2016).

191. Töpf, A. *et al.* Functionally significant, rare transcription factor variants in tetralogy of fallot. *PLoS One* **9**, (2014).

192. Lu, C. X. *et al.* A novel HAND2 loss-of-function mutation responsible for tetralogy of Fallot. *Int. J. Mol. Med.* **37**, 445–451 (2016).

193. Tamura, M. *et al.* Overdosage of Hand2 causes limb and heart defects in the human

chromosomal disorder partial trisomy distal 4q. *Hum. Mol. Genet.* **22**, 2471–2481 (2013).

194. Pasipoularides, A. The new era of whole-exome sequencing in congenital heart disease: brand-new insights into rare pathogenic variants. *J. Thorac. Dis.* **10**, S1923–S1929 (2018).
195. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262–6 (2015).
196. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–3 (2013).
197. Li, G. *et al.* Single cell expression analysis reveals anatomical and cell cycle-dependent transcriptional shifts during heart development. *Development* **146**, (2019).
198. Goodyer, W. R. *et al.* Transcriptomic Profiling of the Developing Cardiac Conduction System at Single-Cell Resolution. *Circ. Res.* (2019). doi:10.1161/CIRCRESAHA.118.314578
199. Cui, Y. *et al.* Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Rep.* **26**, 1934-1950.e5 (2019).
200. Taniguchi, H. *et al.* A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex. *Neuron* **71**, 995–1013 (2011).
201. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
202. Rossant, J., Zirngibl, R., Giguere, V., Cado, D. & Shago, M. Expression of a retinoic acid response element-hsplacZ transgene defines specific domains of

transcriptional activity during mouse embryogenesis. *Genes Dev.* **5**, 1333–1344 (2007).

203. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).

204. Gross-Thebing, T., Paksa, A. & Raz, E. Simultaneous high-resolution detection of multiple transcripts combined with localization of proteins in whole-mount embryos. *BMC Biol.* **12**, 1–13 (2014).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

08/09/2019

Date